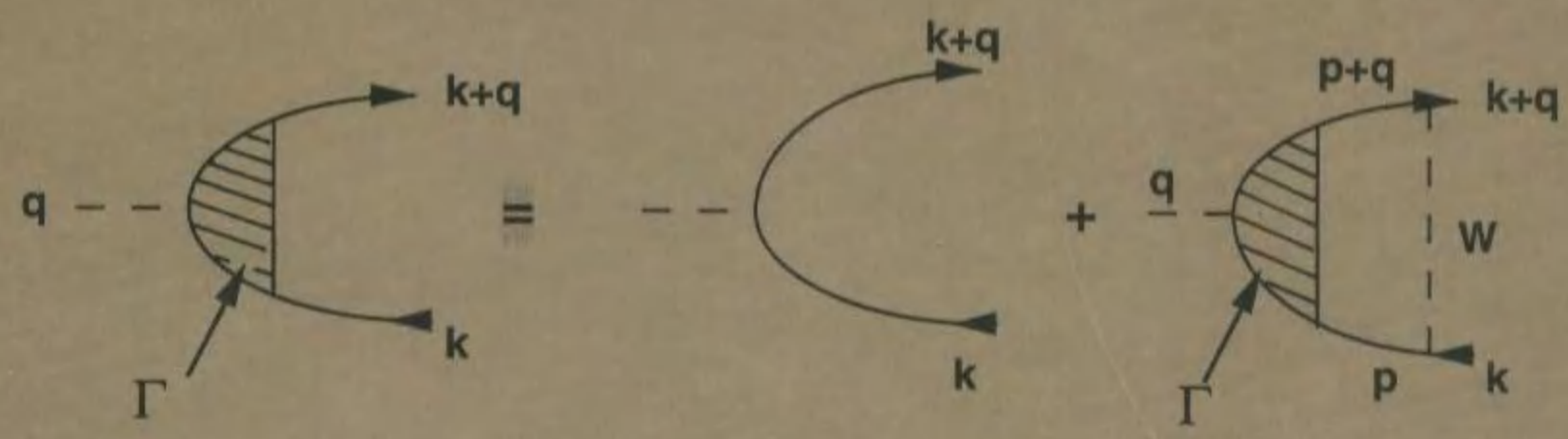
PHYSICS OF SOLIDS AND LIQUIDS

Many-Particle Physics

THIRD EDITION



Gerald D. Mahan

Many-Particle Physics

THIRD EDITION

PHYSICS OF SOLIDS AND LIQUIDS

Editorial Board: **Jozef T. Devreese** • *University of Antwerp, Belgium*
Roger P. Evrary • *University of Liège, Belgium*
Stig Lundqvist • *Chalmers University of Technology, Sweden*
Gerald D. Mahan • *University of Tennessee, USA*
Norman H. March • *University of Oxford, England*

Current Volumes in the Series

CRYSTALLINE SEMICONDUCTING MATERIALS AND DEVICES
Edited by Paul N. Butcher, Norman H. March, and Mario P. Tosi

ELECTRON CORRELATION IN MOLECULES AND CONDENSED PHASES
N. H. March

ELECTRONIC EXCITATIONS AT METAL SURFACES
Ansgar Liebsch

EXCITATION ENERGY TRANSFER PROCESSES IN CONDENSED MATTER: Theory and Applications
Jai Singh

FRACTALS
Jens Feder

INTERACTION OF ATOMS AND MOLECULES WITH SOLID SURFACES
Edited by V. Bortolani, N. H. March, and M. P. Tosi

LOCAL DENSITY THEORY OF POLARIZABILITY
Gerald D. Mahan and K. R. Subbaswamy

MANY-PARTICLE PHYSICS, Third Edition
Gerald D. Mahan

ORDER AND CHAOS IN NONLINEAR PHYSICAL SYSTEMS
Edited by Stig Lungqvist, Norman H. March, and Mario P. Tosi

PHYSICS OF LOW-DIMENSIONAL SEMICONDUCTOR STRUCTURES
Edited by Paul Butcher, Norman H. March, and Mario P. Tosi

QUANTUM TRANSPORT IN SEMICONDUCTORS
Edited by David K. Ferry and Carlo Jacoboni

A Continuation Order Plan is available for this series. A continuation order will bring delivery of each new volume immediately upon publication. Volumes are billed only upon actual shipment. For further information please contact the publisher.

Many-Particle Physics

THIRD EDITION

Gerald D. Mahan

University of Tennessee, Knoxville, Tennessee
and Oak Ridge National Laboratory

Kluwer Academic / Plenum Publishers
New York, Boston, Dordrecht, London, Moscow

Inventar-Nr.: 2001 - B 15

Library of Congress Cataloging-in-Publication Data

Mahan, Gerald D.
Many-particle physics/Gerald D. Mahan.—3rd ed.
p. cm. — (Physics of solids and liquids)
Includes bibliographical references and index.
ISBN 0-306-46338-5
1. Solid state physics. 2. Many-body problem. 3. Green's functions. I. Title. II. Series.

QC176 .M24 2000
530.4'1—dc21

00-039101

ISBN 0-306-46338-5

©2000, 1990, 1981 Kluwer Academic / Plenum Publishers, New York
233 Spring Street, New York, New York 10013

http://www.wkap.nl

10 9 8 7 6 5 4 3 2 1

A C.I.P. record for this book is available from the Library of Congress

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

Printed in the United States of America

Preface

The first, second, and third editions of this book seem to occur at ten year intervals. The intent is to keep the book up-to-date. Many-body theory is a field which continually evolves in time. Journals only publish new results, conferences only invite speakers to report new phenomena, and agencies only fund scientists to do new physics. Today's physics is old hat by tomorrow. Students want to learn new material, and textbooks must be modified to keep up with the times.

The early chapters in this book teach the techniques of many-body theory. They are largely unchanged in format. The later chapters apply the techniques to specific problems. The third edition increases the number of applications. New sections have been added, while old sections have been modified to include recent applications.

The previous editions were set in type using pre-computer technology. No computer file existed of the prior editions. The publisher scanned the second edition and gave me a disk with the contents. This scan recorded the words accurately and scrambled the equations into unintelligible form. So I retyped the equations using LaTeX. Although tedious, it allowed me to correct the infinite numbers of typographical errors in the previous edition. The earlier typesetting methods did not permit such corrections. The entire book was edited sentence-by-sentence. Most old sections of the book were shortened by editing sentences and paragraphs.

I also contemplated removing entirely some old sections. Each time I did this, and told somebody, they always remarked that the deleted section was their favorite, and I simply could not remove it. While it is gratifying to have so many sections be everyone's favorite, it does make shortening the book somewhat hard! In the end I gave up, and no sections were removed. Many were rewritten to shorten them. Since many new sections were added, the book gets longer with each edition. The reference list was updated.

New sections include: Bethe lattice, different mean-free-paths, Hubbard model, Coulomb blockade, Landauer transport, and the Quantum Hall effect. The big problem is what to say about high-temperature superconductivity. Although much experimental information is available regarding this important topic, the theoretical picture is quite uncertain. There is no agreed understanding of the pairing mechanism which causes the high transition temperature. It is hard to write a text book on a topic for which there is little agreement regarding fundamental theory. In the end, I mentioned only some important experiments and their results, and added little new information on the theory mechanisms. The section on the gap equation was rewritten to use the modern method of solving it in complex frequency space, rather than the older method of real frequency space.

I thank Steve Girvin for his proofreading twice the various versions of the section on the Quantum Hall effect, and Koung-An Chao for teaching me about quantum dots. I also very much thank my wife Sally for letting me spend every evening and weekend for one year preparing this new edition.

Contents

1. Introductory Material 1

1.1. Harmonic Oscillators and Phonons 1
1.2. Second Quantization for Particles 11
1.3. Electron–Phonon Interactions 26
1.3.1. Interaction Hamiltonian 27
1.3.2. Localized Electron 29
1.3.3. Deformation Potential 31
1.3.4. Piezoelectric Interaction 32
1.3.5. Polar Coupling 34
1.4. Spin Hamiltonians 36
1.4.1. Homogeneous Spin Systems 38
1.4.2. Impurity Spin Models 43
1.5. Photons 48
1.5.1. Gauges 49
1.5.2. Lagrangian 53
1.5.3. Hamiltonian 55
1.6. Pair Distribution Function 58
Problems 62

2. Green's Functions at Zero Temperature 65

2.1. Interaction Representation 66
2.1.1. Schrödinger 66
2.1.2. Heisenberg 66
2.1.3. Interaction 67
2.2. S Matrix 70
2.3. Green's Functions 71
2.4. Wick's Theorem 76
2.5. Feynman Diagrams 81
2.6. Vacuum Polarization Graphs 83

2.7. Dyson's Equation 86
2.8. Rules for Constructing Diagrams 90
2.9. Time-Loop S Matrix 95
2.9.1. Six Green's Functions 96
2.9.2. Dyson's Equation 99
2.10. Photon Green's Functions 102
Problems 106

3. Nonzero Temperatures 109

3.1. Introduction 109
3.2. Matsubara Green's Functions 112
3.3. Retarded and Advanced Green's Functions 118
3.4. Dyson's Equation 128
3.5. Frequency Summations 136
3.6. Linked Cluster Expansions 142
3.6.1. Thermodynamic Potential 142
3.6.2. Green's Functions 152
3.7. Real-Time Green's Functions 154
3.7.1. Wigner Distribution Function 157
3.8. Kubo Formula for Electrical Conductivity 160
3.8.1. Transverse Fields, Zero Temperature 163
3.8.2. Nonzero Temperatures 168
3.8.3. Zero Frequency 170
3.8.4. Photon Self-Energy 173
3.9. Other Kubo Formulas 174
3.9.1. Pauli Paramagnetic Susceptibility 174
3.9.2. Thermal Currents and Onsager Relations 177
3.9.3. Correlation Functions 181
Problems 183

4. Exactly Solvable Models 187

4.1. Potential Scattering 187
4.1.1. Reaction Matrix 189
4.1.2. T Matrix 192
4.1.3. Friedel's Theorem 195
4.1.4. Impurity Scattering 199
4.1.5. Ground State Energy 204
4.2. Localized State in the Continuum 207
4.3. Independent Boson Models 218
4.3.1. Solution by Canonical Transformation 218
4.3.2. Feynman Disentangling of Operators 221
4.3.3. Einstein Model 224
4.3.4. Optical Absorption and Emission 228
4.3.5. Sudden Switching 236

4.3.6. Linked Cluster Expansion . . . 241
4.4. Bethe Lattice . . . 247
4.4.1. Electron Green's Function . . . 247
4.4.2. Ising Model . . . 251
4.5. Tomonaga Model . . . 256
4.5.1. Tomonaga Model . . . 257
4.5.2. Spin Waves . . . 262
4.5.3. Luttinger Model . . . 264
4.5.4. Single-Particle Properties . . . 267
4.5.5. Interacting System of Spinless Fermions . . . 272
4.6. Polaritons . . . 276
4.6.1. Semiclassical Discussion . . . 276
4.6.2. Phonon–Photon Coupling . . . 278
4.6.3. Exciton–Photon Coupling . . . 282
Problems . . . 291

5. Homogeneous Electron Gas . . . 295

5.1. Exchange and Correlation . . . 295
5.1.1. Kinetic Energy . . . 297
5.1.2. Hartree . . . 297
5.1.3. Exchange . . . 297
5.1.4. Seitz's Theorem . . . 301
5.1.5. $\Sigma^{(2a)}$ . . . 303
5.1.6. $\Sigma^{(2b)}$ . . . 304
5.1.7. $\Sigma^{(2c)}$ . . . 305
5.1.8. High-Density Limit . . . 306
5.1.9. Pair Distribution Function . . . 308
5.2. Wigner Lattice . . . 311
5.3. Metallic Hydrogen . . . 315
5.4. Linear Screening . . . 316
5.5. Model Dielectric Functions . . . 323
5.5.1. Thomas–Fermi . . . 323
5.5.2. Lindhard, or RPA . . . 325
5.5.3. Hubbard . . . 336
5.5.4. Singwi–Sjölander . . . 338
5.5.5. Local Field Corrections . . . 341
5.5.6. Vertex Corrections . . . 343
5.6. Properties of the Electron Gas . . . 346
5.6.1. Pair Distribution Function . . . 346
5.6.2. Screening Charge . . . 346
5.6.3. Correlation Energies . . . 347
5.6.4. Compressibility . . . 352
5.6.5. Pauli Paramagnetic Susceptibility . . . 356
5.7. Sum Rules . . . 358
5.8. One-Electron Properties . . . 362
5.8.1. Renormalization Constant Z_F . . . 365

5.8.2. Effective Mass 368
5.8.3. Mean-Free-Path 369
Problems 372

6. Strong Correlations 375

6.1. Kondo Model 375
6.1.1. High-Temperature Scattering 376
6.1.2. Low-Temperature State 383
6.1.3. Kondo Temperature 387
6.1.4. Kondo Resonance 387
6.2. Single-Site Anderson Model 389
6.2.1. No Hybridization 391
6.2.2. With Hybridization 395
6.2.3. Self-Energy of Electrons 396
6.3. Hubbard Model 403
6.3.1. Spin and Charge Separation 404
6.3.2. Exchange Graphs 409
6.4. Hubbard Model: Magnetic Phases 411
6.4.1. Ferromagnetism 413
6.4.2. Antiferromagnetism 416
6.4.3. An Example 422
6.4.4. Local Field Corrections 427
Problems 430

7. Electron–Phonon Interaction 433

7.1. Fröhlich Hamiltonian 433
7.1.1. Brillouin–Wigner Perturbation Theory 434
7.1.2. Rayleigh–Schrödinger Perturbation Theory 438
7.1.3. Strong Coupling Theory 444
7.1.4. Linked Cluster Theory 448
7.2. Small Polaron Theory 454
7.2.1. Large Polarons 455
7.2.2. Small Polarons 456
7.2.3. Diagonal Transitions 458
7.2.4. Nondiagonal Transitions 459
7.2.5. Kubo Formula 463
7.3. Heavily Doped Semiconductors 467
7.3.1. Screened Interaction 468
7.3.2. Experimental Verifications 474
7.3.3. Electron Self-Energies 475
7.4. Metals 481
7.4.1. Phonons in Metals 482
7.4.2. Electron Self-Energies 487
Problems 495

8. dc Conductivities 499

8.1. Electron Scattering by Impurities 499
8.1.1. Boltzmann Equation 499
8.1.2. Kubo Formula: Approximate Solution 505
8.1.3. Ward Identities 514
8.2. Mobility of Fröhlich Polarons 517
8.3. Electron–Phonon Relaxation Times 524
8.3.1. Metals 526
8.3.2. Semiconductors 527
8.3.3. Temperature Relaxation 531
8.4. Electron–Phonon Interactions in Metals 534
8.4.1. Force–Force Correlation Function 534
8.4.2. Kubo Formula 537
8.4.3. Mass Enhancement 545
8.4.4. Thermoelectric Power 545
8.5. Quantum Boltzmann Equation 549
8.5.1. Derivation of the QBE 550
8.5.2. Gradient Expansion 554
8.5.3. Electron Scattering by Impurities 557
8.6. Quantum Dot Tunneling 561
8.6.1. Electron Tunneling 561
8.6.2. Quantum Dots 567
8.6.3. Rate Equations 571
8.6.4. Quantum Conductance 575
Problems 576

9. Optical Properties of Solids 579

9.1. Nearly Free-Electron Systems 579
9.1.1. General Properties 579
9.1.2. Force–Force Correlation Functions 581
9.1.3. Fröhlich Polarons 585
9.1.4. Interband Transitions 588
9.1.5. Phonons 590
9.2. Wannier Excitons 592
9.2.1. The Model 592
9.2.2. Solution by Green's Functions 596
9.2.3. Core-Level Spectra 600
9.3. X-ray Spectra in Metals 603
9.3.1. Physical Model 603
9.3.2. Edge Singularities 607
9.3.3. Orthogonality Catastrophe 612
9.3.4. MND Theory 621
9.3.5. XPS Spectra 624
Problems 626

10. Superconductivity 627

10.1. Cooper Instability 628
10.1.1. BCS Theory 635
10.2. Superconducting Tunneling 644
10.2.1. Normal–Superconductor 645
10.2.2. Two Superconductors 648
10.2.3. Josephson Tunneling 652
10.2.4. Infrared Absorption 660
10.3. Strong Coupling Theory 664
10.4. Transition Temperature 670
Problems 674

11. Superfluids 677

11.1. Liquid ^{4}He 677
11.1.1. Hartree and Exchange 679
11.1.2. Bogoliubov Theory of ^{4}He 682
11.1.3. Off-Diagonal Long-Range Order 686
11.1.4. Correlated Basis Functions 690
11.1.5. Experiments on $n_{\mathbf{k}}$. 697
11.1.6. Bijl-Feynman Theory 702
11.1.7. Improved Excitation Spectra 707
11.1.8. Superfluidity 710
11.2. Liquid ^{3}He 713
11.2.1. Fermi Liquid Theory 714
11.2.2. Experiments and Microscopic Theories 720
11.2.3. Interaction Between Quasiparticles: Excitations 723
11.2.4. Quasiparticle Transport 729
11.2.5. Superfluid ^{3}He 735
11.3. Quantum Hall Effects 742
11.3.1. Landau Levels 742
11.3.2. Classical Hall Effect 745
11.3.3. Quantum Hall Effect 747
11.3.3.1. Fixed Density 749
11.3.3.2. Fixed Chemical Potential 749
11.3.3.3. Impurity Dominated 750
11.3.4. Laughlin Wave Function. 752
11.3.5. Collective Excitations. 757
11.3.5.1. Magnetorotons 757
11.3.5.2. Quasiholes 760
Problems 761

References 765
Author Index 777
Subject Index 781

Chapter 1

Introductory Material

1.1. HARMONIC OSCILLATORS AND PHONONS

First quantization in physics refers to the property of particles that certain operators do not commute:

$$[x, p_x] = i\hbar \tag{1.1}$$

$$E \rightarrow i\hbar \frac{\partial}{\partial t} \tag{1.2}$$

Later it was realized that forces between particles were caused by other particles: photons caused electromagnetic forces, pions caused some nuclear forces, etc. These particles are also quantized, which leads to second quantization. The basic idea is that forces are caused by the exchange of particles, and the number of particles is quantized: one, two, three, etc. The quantization imparts a quantum nature to the classical force fields.

In solids the vibrational modes of the atoms are quantized because of first quantization (1.1.1). These quantized vibrational modes are called phonons. An electron can interact with a phonon, and this phonon can travel to another electron, interact, and thereby cause an indirect interaction between electrons. Indeed, the phonon need not move but can vibrate until the next electron comes by. The induced interaction between electrons is an example of second quantization. The phonons play a role in solids similar to the classical fields of particle physics. They cause quantized interactions between electrons.

Phonons in solids can usually be described as harmonic oscillators. A fuller description of the effects of anharmonicity is introduced later. But, for the moment, this idea should be sufficient motivation to study the harmonic oscillator. The one-dimensional harmonic oscillator has the Hamiltonian

$$H = \frac{p^2}{2m} + \frac{K}{2}x^2 \tag{1.3}$$

To solve this Hamiltonian, introduce a dimensionless coordinate ξ:

$$\omega^2 = \frac{K}{m} \tag{1.4}$$

$$\xi = x\left(\frac{m\omega}{\hbar}\right)^{1/2} \tag{1.5}$$

$$-i\frac{\partial}{\partial\xi} = \frac{p}{\sqrt{\hbar m\omega}} \tag{1.6}$$

and

$$H = \frac{\hbar\omega}{2}\left(-\frac{\partial^2}{\partial\xi^2} + \xi^2\right) \tag{1.7}$$

The harmonic oscillator Hamiltonian has a solution in terms of Hermite polynomials. The states are quantized such that

$$H\psi_n = \hbar\omega\left(n + \frac{1}{2}\right)\psi_n \tag{1.8}$$

where n is an integer. Use Dirac notation for the eigenstates $|n\rangle = \psi_n$. One can also learn by direct calculation that the following matrix elements exist for the operators x and p:

$$\begin{aligned} \langle n'|x|n\rangle &= \left(\frac{\hbar}{2m\omega}\right)^{1/2}[(n')^{1/2}\delta_{n'=n+1} + (n)^{1/2}\delta_{n'=n-1}] \\ \langle n'|p|n\rangle &= i\left(\frac{m\hbar\omega}{2}\right)^{1/2}[(n')^{1/2}\delta_{n'=n+1} - (n)^{1/2}\delta_{n'=n-1}] \end{aligned} \tag{1.9}$$

It is customary to define two dimensionless operators as follows:

$$\begin{aligned} a &= \frac{1}{\sqrt{2}}\left(\xi + \frac{\partial}{\partial\xi}\right) = \left(\frac{m\omega}{2\hbar}\right)^{1/2}\left(x + \frac{ip}{m\omega}\right) \\ a^\dagger &= \frac{1}{\sqrt{2}}\left(\xi - \frac{\partial}{\partial\xi}\right) = \left(\frac{m\omega}{2\hbar}\right)^{1/2}\left(x - \frac{ip}{m\omega}\right) \end{aligned} \tag{1.10}$$

They are Hermitian conjugates of each other. They are sometimes called *raising* and *lowering operators*, but here they are called *creation* ($a^\dagger$) and *destruction operators* (a). The Hamiltonian (1.7) may be written with them as

$$H = \frac{\hbar\omega}{2}[aa^\dagger + a^\dagger a] \tag{1.11}$$

$$= \frac{\hbar\omega}{2}\left[\frac{1}{2}\left(\xi + \frac{\partial}{\partial\xi}\right)\left(\xi - \frac{\partial}{\partial\xi}\right) + \frac{1}{2}\left(\xi - \frac{\partial}{\partial\xi}\right)\left(\xi + \frac{\partial}{\partial\xi}\right)\right] \tag{1.12}$$

$$H = \frac{\hbar\omega}{2}\left(-\frac{\partial^2}{\partial\xi^2} + \xi^2\right) \tag{1.13}$$

A very important property of these operators is called *commutation relations*. These are derived by considering how they act, sequentially, on any function $f(\xi)$. The two operations a and $a^\dagger$ in turn give

$$aa^\dagger f(\xi) = \frac{1}{2}\left(\xi + \frac{\partial}{\partial\xi}\right)\left(\xi - \frac{\partial}{\partial\xi}\right)f(\xi) = \frac{1}{2}(\xi^2 f + f - f'') \tag{1.14}$$

while the reverse order gives

$$a^\dagger a f(\xi) = \frac{1}{2}\left(\xi - \frac{\partial}{\partial\xi}\right)\left(\xi + \frac{\partial}{\partial\xi}\right)f(\xi) = \frac{1}{2}(\xi^2 f - f - f'') \tag{1.15}$$

These two results are subtracted,

$$[aa^\dagger - a^\dagger a]f(\xi) = f(\xi) \tag{1.16}$$

and yield the original function. The operator in brackets is replaced by a bracket with a comma,

$$[aa^\dagger - a^\dagger a] \equiv [a, a^\dagger] \tag{1.17}$$

which means the same thing. The relationship (1.16) is usually expressed by omitting the function $f(\xi)$:

$$[a, a^\dagger] = 1 \tag{1.18}$$

In a similar way, one can prove that

$$[a, a] = 0 \tag{1.19}$$

$$[a^\dagger, a^\dagger] = 0 \tag{1.20}$$

These three commutators, plus the Hamiltonian

$$H = \frac{\hbar\omega}{2}[aa^\dagger + a^\dagger a] = \frac{\hbar\omega}{2}[aa^\dagger - a^\dagger a + 2a^\dagger a] = \hbar\omega[a^\dagger a + \tfrac{1}{2}] \tag{1.21}$$

completely specify the harmonic oscillator problem in terms of operators. With these four relationships, one can show that the eigenvalue spectrum is indeed (1.8), where n is an integer. The eigenstates are

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle \tag{1.22}$$

where $|0\rangle$ is the state of no phonons which obeys

$$a|0\rangle = 0 \tag{1.23}$$

and where the $n!$ is for normalization. Operating on this state by a creation operator gives

$$a^\dagger|n\rangle = \frac{(a^\dagger)^{n+1}}{\sqrt{n!}}|0\rangle = \frac{(n+1)^{1/2}}{\sqrt{(n+1)!}}(a^\dagger)^{n+1}|0\rangle \tag{1.24}$$

$$= (n+1)^{1/2}|n+1\rangle \tag{1.25}$$

the state with the next highest integer. The only matrix element between states are

$$\langle n'|a^\dagger|n\rangle = (n+1)^{1/2}\delta_{n'=n+1} \tag{1.26}$$

$$\langle n'|a|n\rangle = (n)^{1/2}\delta_{n'=n-1} \tag{1.27}$$

The second expression is derived from the first by taking the Hermitian conjugate of the first, and then exchanging dummy variables n and n'. Alternately,

$$a|n\rangle = (n)^{1/2}|n-1\rangle \tag{1.28}$$

So the destruction operator a lowers the quantum number. Then operating by the sequence

$$a^\dagger a|n\rangle = a^\dagger(n)^{1/2}|n-1\rangle = (n)^{1/2}a^\dagger|n-1\rangle = n|n\rangle \tag{1.29}$$

gives an eigenvalue n, which verifies the eigenvalue (1.8). Furthermore, using the original definitions (1.10) permits us to express x and p in terms of these operators,

$$x = \left(\frac{\hbar}{2m\omega}\right)^{1/2}(a + a^\dagger) \tag{1.30}$$

$$p = i\left(\frac{m\hbar\omega}{2}\right)^{1/2}(a^\dagger - a) \tag{1.31}$$

and the matrix elements (1.9) follow immediately:

$$\begin{aligned}
\langle n'|x|n\rangle &= \left(\frac{\hbar}{2m\omega}\right)^{1/2}[\langle n'|a|n\rangle + \langle n'|a^\dagger|n\rangle] \\
&= \left(\frac{\hbar}{2m\omega}\right)^{1/2}[(n)^{1/2}\delta_{n'=n-1} + (n+1)^{1/2}\delta_{n'=n+1}] \\
\langle n'|p|n\rangle &= i\left(\frac{m\hbar\omega}{2}\right)^{1/2}[\langle n'|a^\dagger|n\rangle - \langle n'|a|n\rangle] \\
&= i\left(\frac{m\hbar\omega}{2}\right)^{1/2}[(n+1)^{1/2}\delta_{n'=n+1} - (n)^{1/2}\delta_{n'=n-1}]
\end{aligned}$$

The description of the harmonic oscillator in terms of operators is equivalent to the conventional method of using wave functions $\psi_n(\xi)$ of position.

The time dependence of these operators is often important. In the Heisenberg representation of quantum mechanics, the time development of operators is given by ($\hbar = 1$)

$$O(t) = e^{iHt}Oe^{-iHt} \tag{1.32}$$

so that the operator obeys the equation

$$i\frac{\partial O(t)}{\partial t} = i[H, O(t)] \tag{1.33}$$

For the destruction operator, this equation becomes

$$\frac{\partial}{\partial t}a = i[H, a] = i\omega[a^\dagger aa - aa^\dagger a] = i\omega[a^\dagger, a]a = -i\omega a \tag{1.34}$$

which has the simple solution

$$a(t) = e^{-i\omega t} a \tag{1.35}$$

The reference point of time may be selected arbitrarily, so that the operators have an arbitrary phase factor associated with them. This phase is unimportant, since it cancels out of all final results. The Hermitian conjugate of this expression is

$$a^\dagger(t) = e^{i\omega t} a^\dagger \tag{1.36}$$

The time development of the position operator can be represented as

$$x(t) = \left(\frac{\hbar}{2m\omega}\right)^{1/2} (a e^{-i\omega t} + a^\dagger e^{i\omega t}) \tag{1.37}$$

This result for $x(t)$ will be used often in discussing phonon problems.

Another familiar problem which can be solved with operators is a charged harmonic oscillator in a constant electric field F:

$$H = \frac{p^2}{2m} + \frac{K}{2}x^2 + eFx = \omega(a^\dagger a + \tfrac{1}{2}) + \lambda(a + a^\dagger) \tag{1.38}$$

$$\lambda = eF\left(\frac{\hbar}{2m\omega}\right)^{1/2} \tag{1.39}$$

This Hamiltonian may be solved exactly. First consider the equation of motion for the time development of the destruction operator:

$$\frac{\partial a}{\partial t} = i[H, a] = -i(\omega a + \lambda) \tag{1.40}$$

The right-hand side is no longer just proportional to a, since there is the constant term. However, let us define a new set of operators by the relationships

$$A = a + \frac{\lambda}{\omega} \tag{1.41}$$

$$A^\dagger = a^\dagger + \frac{\lambda}{\omega} \tag{1.42}$$

They obey the equation

$$\frac{\partial A}{\partial t} = -i\omega A \tag{1.43}$$

so they have the simple time development

$$A(t) = e^{-i\omega t} A \tag{1.44}$$

$$A^\dagger(t) = e^{i\omega t} A^\dagger \tag{1.45}$$

Indeed, one can show that they have the following properties:

$$[A, A^\dagger] = \left[a + \frac{\lambda}{\omega}, a^\dagger + \frac{\lambda}{\omega}\right] = 1 \tag{1.46}$$

$$[A, A] = 0 \tag{1.47}$$

$$[A^\dagger, A^\dagger] = 0 \tag{1.48}$$

$$H = \omega\left[\left(A^\dagger - \frac{\lambda}{\omega}\right)\left(A - \frac{\lambda}{\omega}\right) + \frac{1}{2}\right] + \lambda\left(A + A^\dagger - 2\frac{\lambda}{\omega}\right) \tag{1.49}$$

$$= \omega\left(A^\dagger A + \frac{1}{2}\right) - \frac{\lambda^2}{\omega} \tag{1.50}$$

It was remarked above that any set of operators with these properties had a solution in terms of harmonic oscillator states:

$$H|n\rangle = \left[\omega\left(n + \frac{1}{2}\right) - \frac{\lambda^2}{\omega}\right]|n\rangle \tag{1.51}$$

$$|n\rangle = \frac{(A^\dagger)^n}{\sqrt{n!}}|0\rangle \tag{1.52}$$

The operator for the position is

$$x(t) = \left(\frac{\hbar}{2m\omega}\right)^{1/2}\left(Ae^{-i\omega t} + A^\dagger e^{i\omega t} - \frac{2\lambda}{\omega}\right) \tag{1.53}$$

The physics of the Hamiltonian (1.38) is very simple. The spring stretches to a new equilibrium point which is displaced a distance

$$x_0 = -\left(\frac{\hbar}{2m\omega}\right)^{1/2}\frac{2\lambda}{\omega} = -\frac{eF}{K} \tag{1.54}$$

from the original one. It oscillates about this new equilibrium with the same frequency ω as before. These oscillations are still quantized, in units of ω. The energy $-\lambda^2/\omega = -e^2F^2/2K$ is that gained by the spring from the displacement along the electric field. One can get the same result directly in coordinate space. The Hamiltonian is written as

$$H = \frac{p^2}{2m} + \frac{K}{2}\left(x + \frac{eF}{K}\right)^2 - \frac{e^2F^2}{2K} \tag{1.55}$$

and a new coordinate $x' = x + eF/K$ is defined which obeys

$$[x', p] = i\hbar \tag{1.56}$$

The variable x' describes the simple harmonic motion. This result completes the discussion of the harmonic oscillator in an electric field.

In a solid there are many atoms, which mutually interact. The vibrational modes are collective motions involving many atoms. A simple introduction to this problem is obtained by studying the normal modes of a one-dimensional harmonic chain:

$$H = \sum_i \frac{p_i^2}{2m} + \frac{K}{2}\sum_i (x_i - x_{i+1})^2 \tag{1.57}$$

The classical solution is obtained by solving the equation of motion:

$$-m\ddot{x}_j = m\omega^2 x_j = K(2x_j - x_{j+1} - x_{j-1}) \tag{1.58}$$

A solution is assumed of the form $x_j = x_0 \cos(kaj)$ and the force term becomes

$$\begin{aligned} 2x_j - x_{j+1} - x_{j-1} &= x_0[2\cos(kaj) - \cos(kaj + ka) - \cos(kaj - ka)] \\ &= 2x_0 \cos(kaj)[1 - \cos(ka)] \end{aligned}$$

The normal modes have the solution

$$\omega_k^2 = \frac{2K}{m}[1 - \cos(ka)] = \frac{4K}{m}\sin^2\left(\frac{ka}{2}\right) \tag{1.59}$$

The quantum mechanical solution begins by defining some normal coordinates, assuming periodic boundary conditions:

$$\begin{aligned} x_l &= \frac{1}{\sqrt{N}}\textstyle\sum_k e^{ikal} x_k; \qquad & x_k &= \frac{1}{\sqrt{N}}\textstyle\sum_l e^{-ikal} x_l \\ p_l &= \frac{1}{\sqrt{N}}\textstyle\sum_k e^{-ikal} p_k; \quad & p_k &= \frac{1}{\sqrt{N}}\textstyle\sum_l e^{ikal} p_l \end{aligned} \tag{1.60}$$

This choice maintains the desired commutation relations in either real space or wave vector space:

$$[x_l, p_m] = i\delta_{lm} \tag{1.61}$$

$$[x_k, p_{k'}] = \frac{1}{N}\sum_{l,m} e^{-ikal} e^{ik'am}[x_l, p_m] \tag{1.62}$$

$$= \frac{i}{N}\sum_m e^{iam(k'-k)} = i\delta_{k,k'} \tag{1.63}$$

From the general result

$$\sum_l x_l x_{l+m} = \frac{1}{N}\sum_{kk'} x_k x_{k'} \sum_l e^{ial(k+k')} e^{imak'} = \sum_k x_k x_{-k} e^{iamk}$$

$$\sum_l p_l^2 = \sum_k p_k p_{-k}$$

It is easy to show that the potential energy term in (1.57) is

$$\frac{K}{2}\sum_j (x_j - x_{j+1})^2 = \frac{K}{2}\sum_k x_k x_{-k}(2 - e^{ika} - e^{-ika}) = \frac{m}{2}\sum_k \omega_k^2 x_k x_{-k}$$

The Hamiltonian may be written in wave vector space as

$$H = \frac{1}{2m}\sum_k \left[p_k p_{-k} + m^2\omega_k^2 x_k x_{-k}\right] \tag{1.64}$$

The Hamiltonian has the form of a simple harmonic oscillator for each wave vector. Define the creation and destruction operators as

$$a_k = \left(\frac{m\omega_k}{2\hbar}\right)^{1/2}\left(x_k + \frac{i}{m\omega_k}p_{-k}\right) \tag{1.65}$$

$$a_k^\dagger = \left(\frac{m\omega_k}{2\hbar}\right)^{1/2}\left(x_k - \frac{i}{m\omega_k}p_{-k}\right) \tag{1.66}$$

They obey the commutation relations

$$[a_k, a_{k'}^\dagger] = -\frac{i}{2\hbar}\{[x_k, p_{k'}] - [p_{-k}, x_{-k'}]\} = \delta_{k,k'} \tag{1.67}$$

$$[a_k, a_{k'}] = 0 \tag{1.68}$$

$$[a_k^\dagger, a_{k'}^\dagger] = 0 \tag{1.69}$$

and the Hamiltonian may be written as

$$H = \sum_k \omega_k[a_k^\dagger a_k + \tfrac{1}{2}] \tag{1.70}$$

These collective modes of vibration are called phonons. They are the quantized version of the classical vibrational modes in the solid. These are the same commutator relations, and Hamiltonian, as in the simple harmonic oscillator. Each wave vector state behaves independently, as a harmonic oscillator, with a possible set of quantum numbers $n_k = 0, 1, 2, \ldots$ The state of the system at any time is

$$\psi = |n_1, n_2, \ldots, n_n) = \Pi_k|n_k\rangle = \Pi_k \frac{(a_k^\dagger)^{n_k}}{\sqrt{n_k!}}|0\rangle \tag{1.71}$$

so that the expectation value of the Hamiltonian is

$$H = \sum_k \omega_k(n_k + \tfrac{1}{2}) \tag{1.72}$$

In thermal equilibrium the states have an average value of n_k which is given in terms of the temperature $\beta = 1/k_BT$:

$$\langle n_k\rangle \equiv N_k = \frac{1}{e^{\beta\omega_k} - 1} \equiv n_B(\omega_k) \tag{1.73}$$

The system fluctuates around this average value.

The position operator in wave vector space, and real space, is

$$x_k(t) = \left(\frac{\hbar}{2m\omega_k}\right)^{1/2}(a_k e^{-i\omega_k t} + a_{-k}^\dagger e^{i\omega_k t})$$

$$x_l(t) = \sum_k \left(\frac{\hbar}{2mN\omega_k}\right)^{1/2} e^{ikal}(a_k e^{-i\omega_k t} + a_{-k}^\dagger e^{i\omega_k t}) \tag{1.74}$$

Often mN is replaced by the equivalent quantity $mN = \rho\nu$, where ρ is the mass density and ν is the volume. At some point there is a summation over the discrete set of eigenstates for the system of finite volume ν. It is convenient to change the summation to an integration:

$$\lim_{\nu\to\infty} \frac{1}{\nu} \sum_{\mathbf{k}} f(\mathbf{k}) = \int \frac{d^3k}{(2\pi)^3} f(\mathbf{k}) \tag{1.75}$$

During this change, any delta functions must change from discrete delta functions (called Kronecker deltas) to continuous ones. Since

$$\frac{1}{\nu} f(\mathbf{k}') = \frac{1}{\nu} \sum_{\mathbf{k}} f(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'} = \int \frac{d^3k}{(2\pi)^3} f(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'} \tag{1.76}$$

$$= \frac{(2\pi)^3}{\nu} \int \frac{d^3k}{(2\pi)^3} f(\mathbf{k})\delta(\mathbf{k}-\mathbf{k}') \tag{1.77}$$

it is concluded that

$$\lim_{\nu\to\infty} \delta_{\mathbf{k},\mathbf{k}'} = \frac{(2\pi)^3}{\nu} \delta(\mathbf{k}-\mathbf{k}') \tag{1.78}$$

In general, our preference is to write wave vector summations as discrete summations until it is time to do the integrals and only then make the changes (1.75).

The quantum mechanical solution has the same frequencies as found in the classical solution. Quantum mechanics only enters in a quantization of the amplitude of the oscillation. The phonons occur in discrete numbers with zero, one, two, etc., phonons in each state $\mathbf{k}$. When the average number of phonons is large, $n_k \gg 1$, the quantization is irrelevant, since the system behaves classically. The quantum nature of the field is more important when the average number of phonons in each state $\mathbf{k}$ is small.

In three-dimensional solids, the theory is nearly identical except there are more indices. Suppose there is a potential function between atoms or ions of the form

$$\sum_{ij} V(\mathbf{R}_i - \mathbf{R}_j) \tag{1.79}$$

where $\mathbf{R}_i$ is the position of an atom. If it is vibrating, then denote $\mathbf{R}_i^{(0)}$ as the equilibrium position and $\mathbf{Q}_i$ as the displacement from equilibrium:

$$\mathbf{R}_i = \mathbf{R}_i^{(0)} + \mathbf{Q}_i \tag{1.80}$$

The potential function is expanded in a Taylor series about the equilibrium position:

$$\begin{aligned} V(\mathbf{R}_i - \mathbf{R}_j) = {} & V(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) + (\mathbf{Q}_i - \mathbf{Q}_j)\cdot\nabla V(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) \\ & + \frac{1}{2}(\mathbf{Q}_i - \mathbf{Q}_j)_\mu (\mathbf{Q}_i - \mathbf{Q}_j)_\nu \frac{\partial^2}{\partial R_\mu \partial R_\nu} V(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) \\ & + O(Q^3) \end{aligned} \tag{1.81}$$

The term linear in displacement vanishes, because one defines the equilibrium position $\mathbf{R}^{(0)}$ as the place where the sum of the forces on an ion j is zero:

$$\mathbf{F}_j = \sum_i \nabla V(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) = 0 \tag{1.82}$$

The important term is the one which is quadratic in the displacements. It gives the potential energy of the phonons:

$$V_{ph} = \frac{1}{2}\sum_{ij}(\mathbf{Q}_i - \mathbf{Q}_j)_\mu(\mathbf{Q}_i - \mathbf{Q}_j)_\nu \times \Phi_{\mu\nu}(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) \tag{1.83}$$

$$\Phi_{\mu\nu}(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) = \frac{\partial^2}{\partial R_\mu \partial R_\nu} V(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) \tag{1.84}$$

The interaction is evaluated in wave vector space by trying an expansion of the form

$$\mathbf{Q}_i(t) = i\sum_{\mathbf{k},\lambda}\left(\frac{\hbar}{2MN\omega_{\mathbf{k}\lambda}}\right)^{1/2} \xi_{\mathbf{k},\lambda}(a_{\mathbf{k},\lambda}e^{-\omega_{\mathbf{k}\lambda}t} + a^\dagger_{-\mathbf{k},\lambda}e^{i\omega_{\mathbf{k}\lambda}t})e^{i\mathbf{k}\cdot\mathbf{R}_i^{(0)}} \tag{1.85}$$

where M is the ion mass. The factor of i on the right-hand side of the equation is required to make $\mathbf{Q}_i^\dagger = \mathbf{Q}_i$ since it represents a real displacement in space. Take the Hermitian conjugate of (1.85) and change $\mathbf{k} \to -\mathbf{k}$. Nearly the same result is obtained as found in $\mathbf{Q}_i$. To make it Hermitian requires that

$$\xi^*_{\mathbf{k},\lambda} = -\xi_{-\mathbf{k},\lambda} \tag{1.86}$$

The polarization vectors $\xi_{\mathbf{k},\lambda}$ are assumed to be real but change sign with $\mathbf{k}$ direction, $\xi_{-\mathbf{k}} = -\xi_{\mathbf{k}}$ and the above identity is satisfied. Since the displacement is in three dimensions, there are $3L$ normal modes for each value of wave vector. Here L is the number of atoms per unit cell of the crystal. The index λ runs over these $3L$ values of normal mode. Each mode will have its own eigenfrequency $\omega_{\mathbf{k}\lambda}$. It will also have a polarization vector $\xi_{\mathbf{k},\lambda}$ which specifies the vibrational direction of the ion for each wave vector and mode λ. If there are more than one atom per unit cell, one should add further subscripts to M and $\xi_{\mathbf{k},\lambda}$ to specify the values for each atom per unit cell.

The right-hand side of (1.81) may be written as

$$= \sum_{ij} V(\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}) + \frac{M}{2}\sum_{\mathbf{k},\lambda}\mathbf{Q}_{\mathbf{k}\lambda}\cdot\mathbf{Q}_{-\mathbf{k}\lambda}\omega^2_{\mathbf{k}\lambda} \tag{1.87}$$

The first term is a constant which will be neglected in our discussion of vibrational modes. The eigenvalues $\omega_{\mathbf{k}\lambda}$ are those solved in the harmonic approximation. In this approximation, one retains the quadratic term only in the displacements in the Hamiltonian. To be more careful, a third-order anharmonic term V_3 can be added for the case of one atom per unit cell:

$$H = H_0 + V_3 \tag{1.88}$$

$$H_0 = \frac{1}{2M}\sum_{\mathbf{k},\lambda}(\mathbf{P}_{\mathbf{k}\lambda}\cdot\mathbf{P}_{-\mathbf{k}\lambda} + M^2\omega^2_{\mathbf{k}\lambda}\mathbf{Q}_{\mathbf{k}\lambda}\cdot\mathbf{Q}_{-\mathbf{k}\lambda}) \tag{1.89}$$

$$= \sum_{\mathbf{k},\lambda}\omega_{\mathbf{k}\lambda}(a^\dagger_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda} + \tfrac{1}{2}) \tag{1.90}$$

$$V_3 = \frac{1}{3!}\sum_{ij}(\mathbf{Q}_i - \mathbf{Q}_j)_\mu(\mathbf{Q}_i - \mathbf{Q}_j)_\nu(\mathbf{Q}_i - \mathbf{Q}_j)_\sigma\frac{\partial^3 V}{\partial R_\mu \partial R_\nu \partial R_\sigma} \tag{1.91}$$

and one solves H_0 for the modes. These modes are harmonic oscillator states for each wave vector $\mathbf{k}$ and mode λ. The harmonic approximation applies to any theory of phonons which retains only the terms which are quadratic in the displacements $\mathbf{Q}_j$. Actual solids are described by potential functions which are more complicated than the central force field $V(\mathbf{R}_i - \mathbf{R}_j)$ which we have assumed. For example, in semiconductors there are usually bond bending forces between nearest neighbors. Nevertheless, the Hamiltonian is still written as (1.90) in the harmonic approximation. A complete description of these calculations is given by Born and Huang (1954) or Maradudin *et al.* (1963).

The terms in the Taylor series higher than quadratic are treated as perturbations. These are called anharmonic effects and are very important in solids with light atomic masses: hydrogen, helium, lithium, etc. They are often important in some other solids. The first term is usually cubic in the displacements and has the form shown above. In terms of wave vectors it becomes

$$V_3 = \sum_{\mathbf{k},\mathbf{q}} \sum_{\lambda_1\lambda_2\lambda_3} \mathbf{Q}_{\mathbf{k}\lambda_1} \mathbf{Q}_{\mathbf{q}\lambda_2} \mathbf{Q}_{-\mathbf{k}-\mathbf{q},\lambda_3} M(\mathbf{k}, \mathbf{q})_{\lambda_1\lambda_2\lambda_3} \tag{1.92}$$

The matrix element in the cubic term is quite complicated, and it will not be written out. It is difficult to determine from first principles anyway. If the first two displacements $\mathbf{Q}$ have wave vectors $\mathbf{k}$ and $\mathbf{q}$, the third has $-\mathbf{k} - \mathbf{q}$ to ensure wave vector conservation. This interaction may be written in terms of creation and destruction operators by using (1.85). In this representation, it is apparent that these cubic terms permit one phonon to decay into two and vice versa.

For solids in which the anharmonic terms are important, one must try to include the effects of the cubic perturbation V_3 and perhaps also higher terms such as quadratic. Including these terms is a many-body problem. The effects are quite temperature dependent, so it is necessary to use Green's functions at nonzero temperatures.

A word about notation. We dislike subscripts and superscripts. In discussing phonons, the subscript λ will usually be omitted, although it should be carried in every expression. The summation over phonon modes is really meant to imply summation over wave vectors and modes λ.

1.2. SECOND QUANTIZATION FOR PARTICLES

There are two ways to introduce the subject of creation and destruction operators for particles. The first is to describe their properties and then to omit any proofs. One could just remark that they work, which is why we use them. The second way is to go through elaborate justification arguments. These tend to leave the reader more confused than convinced. Here an intermediate approach is tried. A short justification will be attempted. Our discussion follows Schiff (1968).

The first treatment is for boson particles which cannot be destroyed. It is hard to think of a fundamental particle with this property. The method is usually applied to composite particles such as ^{4}He which contain even numbers of fermions, so that it has boson-like properties. In any case, it is assumed the discussion is for a point particle. If it is in a potential $U(\mathbf{r})$, the one-particle Schrödinger equation is

$$i\hbar\dot{\psi}(\mathbf{r}) = H\psi(\mathbf{r}) = \left[-\frac{\hbar^2\nabla^2}{2m} + U(\mathbf{r})\right]\psi(\mathbf{r}) \tag{1.93}$$

This equation may be derived from the Lagrangian density

$$L = i\hbar\psi^\dagger\dot{\psi} - \frac{\hbar^2}{2m}\nabla\psi^\dagger \cdot \nabla\psi - U(\mathbf{r}, t)\psi^\dagger\psi \tag{1.94}$$

The wave function is complex, with real and imaginary parts. These two parts can be treated as independent variables in the Lagrangian. An alternate procedure is to treat $\psi(\mathbf{r})$ and $\psi^\dagger(\mathbf{r})$ as independent variables. Then the usual variations give

$$\frac{\partial L}{\partial \psi} = -U\psi^\dagger \tag{1.95}$$

$$\frac{\partial L}{\partial \psi_x} = -\frac{\hbar^2}{2m}\psi_x^\dagger \tag{1.96}$$

$$\frac{\partial L}{\partial \dot{\psi}} = i\hbar\psi^\dagger \tag{1.97}$$

When these relations are put into Lagrange's equation,

$$0 = \frac{\partial L}{\partial \psi} - \sum_\mu \frac{\partial}{\partial x_\mu}\left(\frac{\partial L}{\partial(\partial\psi/\partial x_\mu)}\right) - \frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{\psi}} \tag{1.98}$$

$$0 = -U\psi^\dagger + \frac{\hbar^2}{2m}\nabla^2\psi^\dagger - i\hbar\frac{\partial}{\partial t}\psi^\dagger \tag{1.99}$$

the Hermitian conjugate of Schrödinger's equation is recovered. If the same manipulations are tried with $\psi^\dagger$ as the variable, then Schrödinger's equation itself is derived. In the Lagrangian formulation, the momentum which is conjugate to the variable ψ is

$$\pi = \frac{\partial L}{\partial \dot{\psi}} = i\hbar\psi^\dagger(\mathbf{r}) \tag{1.100}$$

The Hamiltonian density is given by

$$\mathcal{H} = \pi\dot{\psi} - L = \frac{\hbar^2}{2m}\nabla\psi^\dagger \cdot \nabla\psi + U\psi^\dagger\psi \tag{1.101}$$

where one integrates over all volume to obtain the Hamiltonian

$$H = \int d^3r\mathcal{H} = \int d^3r\left(-\frac{\hbar^2}{2m}\nabla^2 + U\right)\psi \tag{1.102}$$

where the kinetic energy term was integrated by parts. Since π and ψ are conjugate variables, they obey commutation relations of the form

$$[\psi(\mathbf{r}, t), \pi(\mathbf{r}', t)] = i\hbar\delta(\mathbf{r} - \mathbf{r}') \tag{1.103}$$

or using (1.100),

$$[\psi(\mathbf{r}, t), \psi^\dagger(\mathbf{r}', t)] = \delta(\mathbf{r} - \mathbf{r}') \tag{1.104}$$

A commutation relation of this type is the fundamental basis of second quantization. Although it has been made plausible by the derivation from a Lagrangian, it really is a basic

premise. These commutation relations may be satisfied by introducing creation and destruction operators. Let H have eigenstates and eigenvalues of the form

$$H\phi_\lambda = \varepsilon_\lambda \phi_\lambda \tag{1.105}$$

$$H = -\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) \tag{1.106}$$

The wave function $\psi(\mathbf{r})$ and its conjugate $\psi^\dagger(\mathbf{r})$ are expanded in terms of this basis set:

$$\psi(\mathbf{r}) = \sum_\lambda a_\lambda(t)\phi_\lambda(\mathbf{r}) \tag{1.107}$$

$$\psi^\dagger(\mathbf{r}) = \sum_\lambda a_\lambda^\dagger(t)\phi_\lambda^*(\mathbf{r}) \tag{1.108}$$

The original field commutators (1.104) are satisfied if a and $a^\dagger$ operators have their own commutation relations:

$$[a_\lambda(t), a_{\lambda'}^\dagger(t)] = \delta_{\lambda\lambda'}$$

$$[a_\lambda(t), a_{\lambda'}(t)] = 0$$

$$[a_\lambda^\dagger(t), a_{\lambda'}^\dagger(t)] = 0 \tag{1.109}$$

The commutation relations for the field variables are:

$$[\psi(\mathbf{r}, t), \psi(\mathbf{r}', t)] = 0 \tag{1.110}$$

$$[\psi^\dagger(\mathbf{r}, t), \psi^\dagger(\mathbf{r}', t)] = 0 \tag{1.111}$$

$$[\psi(\mathbf{r}, t), \psi^\dagger(\mathbf{r}', t)] = \sum_{\lambda\lambda'}[a_\lambda(t), a_{\lambda'}^\dagger(t)]\phi_\lambda(\mathbf{r})\phi_{\lambda'}^*(\mathbf{r}') \tag{1.112}$$

$$= \sum_\lambda \phi_\lambda(\mathbf{r})\phi_\lambda^*(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \tag{1.113}$$

One might also ask about the commutation relations at different times. How does one evaluate

$$[a_\lambda(t), a_{\lambda'}^\dagger(t')] = \ ? \tag{1.114}$$

The answer is that evaluating this expression is a many-body problem. In fact, that is one of the goals of Green's function theory. The commutator at different times is related to the retarded Green's function, which is defined and discussed in Sec. 3.3. The commutator at different times is one property of the time development of the many-body system. Simple commutation relations such as (1.109) are valid only if the operators are at the same time.

The Hamiltonian is

$$H = \int d^3r\psi^\dagger(\mathbf{r})\mathcal{H}\psi(\mathbf{r}) = \sum_{\lambda\lambda'} a_\lambda^\dagger a_{\lambda'} \int d^3r\phi_\lambda^*(\mathbf{r})\mathcal{H}\phi_{\lambda'}(\mathbf{r}') = \sum_\lambda \varepsilon_\lambda a_\lambda^\dagger a_\lambda \tag{1.115}$$

In Sec. 1.1 it was noted that any system with these commutation relations, and a Hamiltonian (1.115), behaved as harmonic oscillators for each state λ. The eigenstates for each value of λ have a discrete set of occupation numbers $n_\lambda = 0, 1, 2, 3, \ldots$. All bosons have harmonic oscillator eigenstates. For phonons, the number n_λ is interpreted as the number of phonons in state λ. For particles, the interpretation is the same. The number n_λ tells how many particles in

the system are in the same state λ. However, for particles, unlike phonons, the total number of particles is conserved. The many-particle wave function has the form

$$\Pi_\lambda \frac{(a_\lambda^\dagger)^{n_\lambda}}{\sqrt{n_\lambda!}} |0\rangle \tag{1.116}$$

so that the Hamiltonian (1.115) has the eigenvalue of

$$E = \sum_\lambda \varepsilon_\lambda n_\lambda \tag{1.117}$$

In thermal equilibrium, the average number of particles in a state λ is given by the usual boson occupation factor:

$$\langle n_\lambda \rangle = \frac{1}{e^{\beta(\varepsilon_\lambda - \mu)} - 1} \equiv N_\lambda = n_B(\varepsilon_\lambda - \mu) \tag{1.118}$$

Now there is a chemical potential, μ, which can vary with temperature and concentration. It is absent in the phonon, and photon, cases because these excitations do not conserve particle number. One may make as many phonons or photons as one wishes. Another operator of interest is the density operator:

$$\rho(\mathbf{r}) = \psi^\dagger(\mathbf{r})\psi(\mathbf{r}) = \sum_{\lambda\lambda'} a_\lambda^\dagger a_{\lambda'} \phi_\lambda^*(\mathbf{r})\phi_{\lambda'}(\mathbf{r}) \tag{1.119}$$

The integral of $\rho(\mathbf{r})$ is just the number operator:

$$N = \int d^3r \rho(\mathbf{r}) = \sum_\lambda a_\lambda^\dagger a_\lambda \tag{1.120}$$

Its thermal average is obtained simply by taking the thermal average of n_λ:

$$\langle N \rangle = \sum_\lambda \langle n_\lambda \rangle = \sum_\lambda N_\lambda \tag{1.121}$$

This equation serves as a definition of the chemical potential and determines its variations with temperature and particle number N.

The Hamiltonian (1.115) and number operator (1.120) are bilinear in creation and destruction operators. They contain only two operators, one of each kind. Hamiltonians of these kinds may always be solved, at least in principle. The problem may always be reduced to the diagonalization of a matrix. For example, consider the solution of our Hamiltonian:

$$H = -\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) \tag{1.122}$$

Suppose that we were unable to solve it exactly—which is an improbable assumption, since Schrödinger's equation for one-particle potentials may be solved in milliseconds on the computer. Anyway, suppose there were another complete set of states ϕ_n which are the solution to some other Hamiltonian. Expand the wave function in terms of these states,

$$\psi(\mathbf{r}) = \sum_n b_n \phi_n(\mathbf{r}) \tag{1.123}$$

where the creation and destruction operators b_n have the usual commutation relations:

$$[b_n, b_m^\dagger] = \delta_{n,m} \tag{1.124}$$

$$[b_n, b_m] = 0 \tag{1.125}$$

$$[b_n^\dagger, b_m^\dagger] = 0 \tag{1.126}$$

For the Hamiltonian and number operators:

$$H = \sum_{nm} b_n^\dagger b_m H_{nm} \tag{1.127}$$

$$N = \sum_n b_n^\dagger b_n \tag{1.128}$$

$$H_{nm} = \int d^3 r \phi_n^*(\mathbf{r}) H \phi_m(\mathbf{r}) \tag{1.129}$$

This Hamiltonian may be solved in the following fashion. Examine the equation of motion for the destruction operator:

$$-i\frac{\partial}{\partial t} b_n = [H, b_n] = -\sum_m H_{nm} b_m \tag{1.130}$$

It is assumed that an operator has the time development

$$b_n(t) = b_n(0) e^{-iEt} \tag{1.131}$$

so that the solutions are of the form

$$0 = \sum_m (H_{nm} - E\delta_{nm}) \tag{1.132}$$

The eigenvalues E are the solution to

$$\det|H_{nm} - E\delta_{nm}| = 0 \tag{1.133}$$

It is only necessary to find the eigenvalues of the Hamiltonian matrix. Usually the matrix is of infinite dimensionality, since there are an infinite number of states in the set ϕ_n. But one may often diagonalize it exactly for many problems. Computers allow very accurate solutions for any case of interest. If all Hamiltonians had only bilinear operators, then many-body theory would only be an exercise in matrix diagonalization. Fortunately, it is more fun than that.

Many-body theory is used to study Hamiltonians which have additional terms. These terms may be interactions with phonons, spin effects, or particle–particle interactions. The effects of particle–particle interactions may be understood by examining a many-particle Hamiltonian of the form

$$H = \sum_i H_i + \frac{1}{2}\sum_{i \neq j} V(\mathbf{r}_i - \mathbf{r}_j) \tag{1.134}$$

$$H_i = -\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) \tag{1.135}$$

The first term contains a summation of one-particle Hamiltonians H_i. This term by itself is just as simple to solve as H_i alone. A collection of particles which do not mutually interact makes for a trivial problem. One solves the dynamics of one particle, and the total properties are the summation of the individual ones. The term which makes the Hamiltonian hard to

solve is the particle–particle interaction $V(\mathbf{r}_i - \mathbf{r}_j)$. This term is multiplied by one-half since the double summation over (ij) counts each pair twice. This interaction is written in terms of creation and destruction operators:

$$H = \sum_{nm} H_{nm} b_n^\dagger b_m + \tfrac{1}{2} \sum_{klmn} V_{klmn} b_k^\dagger b_m^\dagger b_n b_l$$

$$V_{klmn} = \int d^3 r_i \int d^3 r_j \phi_k^*(\mathbf{r}_i) \phi_l(\mathbf{r}_i) V(\mathbf{r}_i - \mathbf{r}_j) \phi_m^*(\mathbf{r}_j) \phi_n(\mathbf{r}_j) \tag{1.136}$$

The interaction term contains two creation and two destruction particles. This term is interpreted as describing two-particle scattering events. One particle in state l scatters to state k, while another in state n scatters to state m. The process is illustrated in Fig. 1.1. Each index (k, l, m, n) runs over all possible values. There are processes where, for example, $k = l$, which describes a process where one particle scatters from n to m, while the other does not change its state. This term was carefully written so that both destruction operators are to the right and the creation operators to the left. The reason for this arrangement is to eliminate processes whereby a particle interacts with itself. For example, if this expression is written as

$$b_k^\dagger b_l b_m^\dagger b_n \tag{1.137}$$

then we could have $m = l$, and this term would describe how one particle interacts with itself. For example, if $|0\rangle$ is the particle vacuum where all $n_m = 0$, then $b_k|0\rangle = 0$. The state $b_\alpha^\dagger|0\rangle$ contains one particle in state α. The interaction operator

$$\tfrac{1}{2} \sum_{klmn} V_{klmn} b_k^\dagger b_m^\dagger b_n b_l b_\alpha^\dagger |0\rangle = 0 \tag{1.138}$$

on this state gives zero, since two particles cannot interact if there is only one particle in the system. However, the incorrect version of the interaction gives

$$\tfrac{1}{2} \sum_{klmn} V_{klmn} b_k^\dagger b_l b_m^\dagger b_n b_\alpha^\dagger |0\rangle = \tfrac{1}{2} \sum_{km} V_{kmm\alpha} b_k^\dagger |0\rangle \tag{1.139}$$

is nonzero, and this term ($\alpha = n, l = m$) must just be the particle interacting with itself. These terms are avoided by writing the pairwise interaction in the form (1.136).

For a gas or liquid of ^{4}He particles, a common basis set is just free-particle wave functions,

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} a_{\mathbf{k}} \tag{1.140}$$

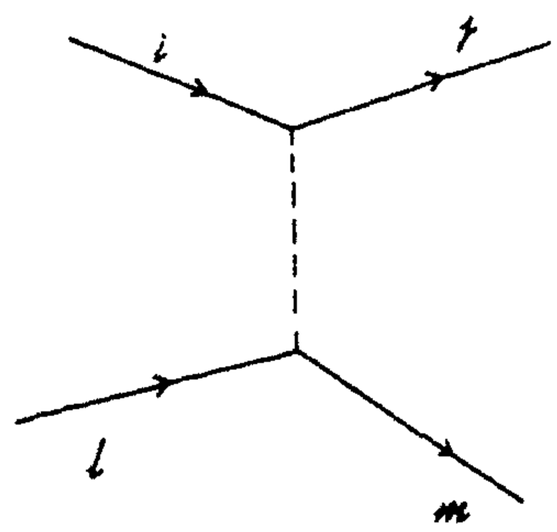


FIGURE 1.1

in which case the Hamiltonian has the form

$$H = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} a_{\mathbf{k}}^{\dagger} a_{\mathbf{k}} + \frac{1}{2\nu} \sum_{\mathbf{k}\mathbf{k}'\mathbf{q}} V(\mathbf{q}) a_{\mathbf{k}+\mathbf{q}}^{\dagger} a_{\mathbf{k}'-\mathbf{q}}^{\dagger} a_{\mathbf{k}'} a_{\mathbf{k}}$$

$$\varepsilon_{\mathbf{k}} = \frac{k^2}{2m}$$

$$V(\mathbf{q}) = \int d^3 r e^{i\mathbf{q}\cdot\mathbf{r}} V(\mathbf{r}) \tag{1.141}$$

The operator $\rho_{\mathbf{q}}$ is the particle density operator in the plane-wave representation:

$$\rho_{\mathbf{q}} = \sum_{\mathbf{k}} a_{\mathbf{k}+\mathbf{q}}^{\dagger} a_{\mathbf{k}} \tag{1.142}$$

The simplest way to write the interaction term in (1.141) is in the form

$$\frac{1}{2\nu} \sum_{\mathbf{q}} V(\mathbf{q}) \rho_{\mathbf{q}} \rho_{-\mathbf{q}} \tag{1.143}$$

which is defective because both destruction operators are not to the right of the creation operators. There is a term where the particle interacts with itself.

A possible difficulty with (1.141) is that $V(\mathbf{q})$ may not exist. The potential $V(\mathbf{r})$ may not possess a Fourier transform if the particle–particle potential is too divergent at small values of $\mathbf{r}$. This divergence happens, for example, with the Lennard-Jones potential, which is often used to represent the helium–helium potential. This difficulty may be avoided by summing subsets of diagrams to get a T-matrix interaction, which is always well behaved.

So far the discussion has concerned boson operators and boson Hamiltonians. Now consider fermions. These are usually electrons, although occasionally one studies holes, positrons, or ^{3}He particles. Fermions have the property that any state may contain only zero or one particle, which is the famous exclusion principle. Jordan and Wigner (1928) discovered that Fermi statistics could be accomplished by making the fields anticommute, which is represented by curly brackets:

$$\psi(\mathbf{r})\psi^{\dagger}(\mathbf{r}') + \psi^{\dagger}(\mathbf{r}')\psi(\mathbf{r}) \equiv \{\psi(\mathbf{r}), \psi^{\dagger}(\mathbf{r}')\} = \delta(\mathbf{r} - \mathbf{r}') \tag{1.144}$$

$$\{\psi(\mathbf{r}), \psi(\mathbf{r}')\} = 0 \tag{1.145}$$

$$\{\psi^{\dagger}(\mathbf{r}), \psi^{\dagger}(\mathbf{r}')\} = 0 \tag{1.146}$$

If these wave functions are expanded in a basis set $\phi_{\lambda}(\mathbf{r})$,

$$\psi(\mathbf{r}) = \sum_{\lambda} c_{\lambda} \phi_{\lambda}(\mathbf{r}) \tag{1.147}$$

$$\psi^{\dagger}(\mathbf{r}) = \sum_{\lambda} c_{\lambda}^{\dagger} \phi_{\lambda}^{*}(\mathbf{r}) \tag{1.148}$$

the coefficients $c_\lambda^\dagger$ and c_λ become creation and destruction operators which obey anti-commutation relations:

$$\{c_\lambda, c_{\lambda'}^\dagger\} = \delta_{\lambda\lambda'} \tag{1.149}$$

$$\{c_\lambda, c_{\lambda'}\} = 0 \tag{1.150}$$

$$\{c_\lambda^\dagger, c_{\lambda'}^\dagger\} = 0 \tag{1.151}$$

For example, consider $\{c_\lambda, c_\lambda\} = 2c_\lambda c_\lambda = 0$. The operator $c_\lambda c_\lambda$ acting upon anything gives zero, since c_λ is a destruction operator which destroys a particle from a state λ. A state may only contain zero or one particle, which are called $|0\rangle_\lambda$ and $|1\rangle_\lambda$, respectively:

$$\begin{aligned} c_\lambda|1\rangle_\lambda &= |0\rangle_\lambda \quad & c_\lambda^\dagger|1\rangle_\lambda &= 0 \\ c_\lambda|0\rangle_\lambda &= 0 \quad & c_\lambda^\dagger|0\rangle_\lambda &= |1\rangle_\lambda \end{aligned} \tag{1.152}$$

So $c_\lambda c_\lambda$ acting upon either $|1\rangle_\lambda$ or $|0\rangle_\lambda$ gives zero. Similarly, the combination $c_\lambda^\dagger c_\lambda^\dagger = 0$. It is zero because two particles cannot be created in the same state. Another way to see this result is to consider the number operator for a state,

$$N_\lambda = c_\lambda^\dagger c_\lambda \tag{1.153}$$

and its square,

$$N_\lambda^2 = c_\lambda^\dagger c_\lambda c_\lambda^\dagger c_\lambda \tag{1.154}$$

Using the anticommutation relations $cc^\dagger = 1 - c^\dagger c$ gives

$$N_\lambda^2 = c_\lambda^\dagger(1 - c_\lambda^\dagger c_\lambda)\, c_\lambda = c_\lambda^\dagger c_\lambda - c_\lambda^\dagger c_\lambda^\dagger c_\lambda c_\lambda = N_\lambda \tag{1.155}$$

The only numbers which are equal to its square are 0 and 1. The number N_λ may only be 0 or 1. The anticommutation relations have built into them the fermion property that no more than one particle may be in the same state λ.

Quite often the Hamiltonians for fermion calculations are of the form

$$H = \sum_i \left[\frac{p_i^2}{2m} + U(\mathbf{r}_i) + \tfrac{1}{2}\sum_{j\neq i} V(\mathbf{r}_i - \mathbf{r}_j) \right] \tag{1.156}$$

with the particles interacting with a potential $U(\mathbf{r})$ and with each other through particle–particle interactions $V(\mathbf{r}_i - \mathbf{r}_j)$. The Hamiltonian is written in terms of creation and destruction operators exactly as in the boson case:

$$\begin{aligned} H &= \sum_{nm} H_{nm} c_n^\dagger c_m + \tfrac{1}{2}\sum_{klmn} V_{klmn} c_k^\dagger c_m^\dagger c_n c_l \\ H_{nm} &= \int d^3r \phi_n(\mathbf{r})^* \left[-\frac{\hbar^2\nabla^2}{2m} + U(\mathbf{r}) \right] \phi_m(\mathbf{r}) \\ V_{klmn} &= \int d^3r_i \int d^3r_j \phi_k^*(\mathbf{r}_i)\phi_l(\mathbf{r}_i) V(\mathbf{r}_i - \mathbf{r}_j) \phi_m^*(\mathbf{r}_j)\phi_n(\mathbf{r}_j) \end{aligned} \tag{1.157}$$

The difference in behavior between fermions and bosons is often due to the difference in the commutation relations of the operators. The starting Hamiltonians are of similar form, e.g., for liquid ^{3}He and ^{4}He.

The particle–particle interaction term is still interpreted in terms of Fig. 1.1. Great care must be used in writing this term, or else one makes a sign mistake. For example, the term

$$\frac{1}{2}\sum_{klmn} V_{klmn} c_k^\dagger c_m^\dagger c_m^\dagger c_l c_n \tag{1.158}$$

is wrong. The order of the two destruction operators was exchanged. But because of the anticommutation relation

$$c_l c_n = -c_n c_l \tag{1.159}$$

this interchange causes a sign change. Equation (1.158) would be correct if the entire term were multiplied by -1.

For the study of electrons in solids, a popular basis set is plane waves. Eigenstates are described by $(\mathbf{p}, \sigma)$, where σ is the spin index, which is ± 1 for spin up or down. The Hamiltonian then has the form

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} c_{\mathbf{p}\sigma}^\dagger c_{\mathbf{p}\sigma} + \sum_{\mathbf{q}} U(\mathbf{q})\rho_{\mathbf{q}} + \frac{1}{2\mathcal{V}} \sum_{\mathbf{q}\mathbf{k}\mathbf{k}'\sigma\sigma'} v_{\mathbf{q}} c_{\mathbf{k}+\mathbf{q}\sigma}^\dagger c_{\mathbf{k}'-\mathbf{q}\sigma'}^\dagger c_{\mathbf{k}'\sigma'} c_{\mathbf{k}\sigma} \tag{1.160}$$

The second term represents the interaction between the electrons and the atoms or ions of the solid. The interaction $U(\mathbf{q})$ is often represented by a pseudopotential (Harrison, 1966; Heine, 1970).

The electron density operator

$$\rho(\mathbf{q}) = \sum_{\mathbf{k}\sigma} c_{\mathbf{k}+\mathbf{q}\sigma}^\dagger c_{\mathbf{k}\sigma} \tag{1.161}$$

is the same as for bosons except for the additional summation over the spin index. The last term in (1.160) contains the electron–electron interaction, which is just a Coulomb potential. The Fourier transform of a Coulomb potential e^2/r occurs often and has been given a special symbol with lowercase v_q. The Fourier transform is

$$\begin{aligned} v_q &= e^2 \int \frac{d^3 r}{r} e^{i\mathbf{q}\cdot\mathbf{r}} = 2\pi e^2 \int_0^\infty r dr \int_{-1}^{1} d(\cos\theta)\, e^{iqr\cos\theta} \\ &= \frac{2\pi e^2}{iq} \int_0^\infty dr (e^{iqr} - e^{-iqr}) = \frac{4\pi e^2}{q} \int_0^\infty dr \sin(qr) \\ &= \frac{4\pi e^2}{q^2}\left(1 - \lim_{r\to\infty} \cos(qr)\right) \end{aligned} \tag{1.162}$$

The integral is not well defined, since it oscillates at infinity. It is assumed these oscillations damp out, so that the result is

$$v_q = \frac{4\pi e^2}{q^2} \tag{1.163}$$

The electron–electron interactions are a significant part of many-body theory. Most calculations involve worrying about electron–electron interactions in one form or another.

The full electron gas Hamiltonian (1.160) is often too complicated to use for the more elaborate many-body theory. Quite often it is approximated by a model Hamiltonian which has a simpler form. Usually these model Hamiltonians look very simple but still are impossible to solve exactly. Often they are even difficult to solve approximately! Some of these popular models are discussed next.

The *homogeneous electron gas* is a model which is studied frequently to learn about correlation effects. It has the Hamiltonian

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} c^{\dagger}_{\mathbf{p}\sigma} c_{\mathbf{p}\sigma} + \frac{1}{2\nu} \sum_{\mathbf{q}\neq 0, \mathbf{k}\mathbf{k}'\sigma\sigma'} v_q c^{\dagger}_{\mathbf{k}+\mathbf{q}\sigma} c^{\dagger}_{\mathbf{k}'-\mathbf{q}\sigma'} c_{\mathbf{k}'\sigma'} c_{\mathbf{k}\sigma} \tag{1.164}$$

The basic premise is to get rid of the atoms and to replace them with a uniform positive background charge of density n_0. The homogeneous electron gas is also called the *jellium* model. One can think of taking the positive charge of the ions and spreading it uniformly about the unit cell of the crystal. Of course, the homogeneous electron gas then has no crystal structure. To preserve charge neutrality, the average particle density of the electron gas must also be n_0. The average density of the electrons is just the $\mathbf{q} = 0$ value of the density operator,

$$n_0 = \frac{1}{\nu} \langle \rho(q = 0) \rangle = \frac{1}{\nu} \sum_{\mathbf{p}\sigma} \langle c^{\dagger}_{\mathbf{p}\sigma} c_{\mathbf{p}\sigma} \rangle = \frac{1}{\nu} \sum_{\mathbf{p}\sigma} N_{\mathbf{p}} = \frac{N_e}{\nu} \tag{1.165}$$

since the number operator may be summed to give the number of particles N_e. In writing the Hamiltonian (1.164), the $\mathbf{q} = 0$ term in the interaction term was omitted from the summation. The omitted term has the form

$$\frac{N_e(N_e - 1)}{2\nu} \lim_{\mathbf{q}\to 0} v_q \tag{1.166}$$

This term was canceled by two other terms. One of these is the Coulomb interaction of the uniform positive background with itself:

$$\tfrac{1}{2} e^2 n_0^2 \int \frac{d^3r d^3r'}{|\mathbf{r} - \mathbf{r}'|} = \frac{N_e^2}{2\nu} \lim_{\mathbf{q}\to 0} v_q \tag{1.167}$$

The other is the Coulomb interaction of the uniform positive background with the electrons:

$$-e^2 n_0 \int \rho(\mathbf{r}) \frac{d^3r d^3r'}{|\mathbf{r} - \mathbf{r}'|} = -\frac{N_e^2}{\nu} \lim_{\mathbf{q}\to 0} v_q \tag{1.168}$$

The N_e^2 term cancels when these three terms are added, and the other term may be neglected. The Hamiltonian (1.164) describes a system which has charge neutrality. The interaction terms with $\mathbf{q} \neq 0$ describe the fluctuations which occur because of electrons interacting with each other.

The plane-wave model is often a poor approximation of electron behavior in ionic solids. In many solids the electrons are localized on atomic sites and only occasionally hop to neighboring sites. This behavior is described by the *tight-binding* model. One simple form of this model is bilinear in the operators:

$$H = \sum_{j\delta\sigma} W_\delta c^{\dagger}_{j+\delta,\sigma} c_{j\sigma} \tag{1.169}$$

The index j denotes a site at point $\mathbf{R}_j$, while $j + \delta$ represents the nearest neighbor atoms. The number of nearest neighbors is called the *coordination number* and is represented by the symbol $\mathcal{Z}$. One can think of the term

$$W_\delta = \int d^3r \phi^*(\mathbf{r} - \mathbf{R}_j) \left[-\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) \right] \phi(\mathbf{r} - \mathbf{R}_{j+\delta}) \tag{1.170}$$

as arising from the matrix elements between the orbitals $\phi(\mathbf{r})$ which are localized on sites $\mathbf{R}_j$ and $\mathbf{R}_{j+\delta}$. The term W_δ for $\delta \neq 0$ represents processes where the electron jumps from site j to

$j+\delta$, while W_0 is the site energy. The potential $U(\mathbf{r})$ is periodic. Simple versions of the model usually have only a single orbital state for each atomic site. More realistic versions of the tight binding model allow for the multiple orbitals characteristic of p- or d-electrons. The discussion will assume a single orbital state per atomic site.

The bilinear form of the Hamiltonian is trivial to solve exactly. One defines the operator in wave vector space in the usual fashion,

$$c_{j\sigma} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_j} c_{\mathbf{k}\sigma} \tag{1.171}$$

$$c_{\mathbf{k}\sigma} = \frac{1}{\sqrt{N}} \sum_{j} e^{-i\mathbf{k}\cdot\mathbf{R}_j} c_{j\sigma} \tag{1.172}$$

and the Hamiltonian may be rewritten as

$$H = \sum_{\mathbf{k}\sigma} W(\mathbf{k}) c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} \tag{1.173}$$

$$W(\mathbf{k}) = \sum_{\delta} e^{i\mathbf{k}\cdot\delta} W_{\delta} \tag{1.174}$$

If the $c_{j\sigma}$ are fermion operators, then so are the $c_{\mathbf{k}\sigma}$. They obey anticommutation relations $\{c_{\mathbf{k}\sigma}, c^{\dagger}_{\mathbf{k}'\sigma'}\} = \delta_{\mathbf{k}=\mathbf{k}'}\delta_{\sigma\sigma'}$. Each mode $(\mathbf{k}\sigma)$ becomes an independent Fermi system, which may be treated separately in thermodynamic averages. The particle energy is $W(\mathbf{k})$. A very common model is the nearest neighbor model. Here the hopping term is limited to just the nearest neighbor ions, which are presumed to be all alike. Usually the site energy is W_0 set equal to zero, which is just an arbitrary energy renormalization, and $W_\delta = w$. Then the energy is $W(\mathbf{k}) = w\mathcal{Z}\gamma_{\mathbf{k}}$, where the factor

$$\mathcal{Z}\gamma_{\mathbf{k}} = \sum_{\delta} e^{i\mathbf{k}\cdot\delta} \tag{1.175}$$

is summed over the $\mathcal{Z}$ nearest neighbors. This solution is exact for the tight-binding model. For example, the partition function Z_p for fermions is written as

$$Z_p = e^{-\beta\Omega} = \mathrm{Tr}\, e^{-\beta(H-\mu N)} = \Pi_{\mathbf{k}\sigma}\{1 + e^{\beta[W(\mathbf{k})-\mu]}\} \tag{1.176}$$

where Tr denotes the trace. The many-particle partition function (1.176) is obtained by treating each state $\mathbf{k}$ as independent and averaging its thermodynamic properties separately. It appears as if each electron in each state $\mathbf{k}$ is behaving independently. This picture is deceptive. The electrons are not just whizzing around independently.

For example, if one were to calculate the probability that any two electrons of the same spin are on the same atomic site, a zero answer is obtained. Two electrons of the same spin are never on the same site in the single orbital models since they would be in the same state. Of course, it is a basic feature of fermion many-particle wave functions that two electrons of the same spin can never be in the same location. The motion of the electrons is correlated, which arises from the antisymmetry of the wave function under exchange of particle position. The motion of the electrons is not really independent.

This correlation is built into the eigenstates but does not affect the energy in the simple bilinear model, which ignores interactions between particles. The partition function (1.176) has the appearance of independent particle form, even with correlation in the wave function. The antisymmetrization of the many-particle wave function does not affect the expectation

value of one-particle operators. It does affect the expectation value of two (or more)-particle operators.

The tight-binding Hamiltonian may also contain the Coulomb interaction between electrons. In its most general form (1.157), the interaction term is

$$\frac{1}{2}\sum_{klmn} V_{klmn} c_k^\dagger c_m^\dagger c_n c_l \tag{1.177}$$

$$V_{klmn} = \int d^3r_1 \int d^3r_2 \phi^*(\mathbf{r}_1 - \mathbf{R}_k)\phi(\mathbf{r}_1 - \mathbf{R}_l)\frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}$$
$$\times \phi^*(\mathbf{r}_2 - \mathbf{R}_m)\phi(\mathbf{r}_2 - \mathbf{R}_n) \tag{1.178}$$

The four orbitals could be centered on four different sites. These are called four-center integrals. They are usually small and nearly always neglected in many-body calculations. The terms which are included are just the largest Coulomb terms. One possible term is the direct interaction between two particles on different atomic sites. For example, setting $k = l$ and $n = m$ for $k \neq m$ gives

$$\frac{1}{2}\sum_{k\neq m} V_{km} n_k n_m \tag{1.179}$$

$$V_{km} = \int d^3r_1 \int d^3r_2 |\phi(\mathbf{r}_1 - \mathbf{R}_k)|^2 \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} |\phi^*(\mathbf{r}_2 - \mathbf{R}_m)|^2 \tag{1.180}$$

Note the operators can be rearranged into $n_k n_m$ as long as $k \neq m$, so that one does not have a particle interacting with itself. At large separation the interaction terms V_{km} become just a Coulomb potential,

$$\lim_{|R_k - R_m| \to \infty} V_{km} \to \frac{e^2}{|\mathbf{R}_k - \mathbf{R}_m|} \tag{1.181}$$

if the orbitals have s-symmetry. A Hamiltonian with these nearest neighbor Coulomb interactions is similar to the lattice gas Hamiltonians introduced in Sec. 1.4.

The *Hubbard model* (1963) retains only the Coulomb integral which is the very largest. All four orbitals $\phi(\mathbf{r})$ are centered on the same site m. This term describes the interaction between two electrons which are on the same atom. Since two electrons cannot be in the same state, the two on the same atom must be in different atomic states. In the simplest model, which considers only a single orbital state on each atom, the two electrons must have different spin configurations. One has spin up, while the other has spin down. The Hubbard model considers the following Hamiltonian:

$$H = w\sum_{j\delta\sigma} c_{j+\delta,\sigma}^\dagger c_{j\sigma} + W_0 \sum_{j\sigma} c_{j\sigma}^\dagger c_{j\sigma} + U\sum_j n_{j\uparrow} n_{j\downarrow} \tag{1.182}$$

$$U = V_{kk} = \int d^3r_1 \int d^3r_2 |\phi(\mathbf{r}_1)|^2 \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} |\phi(\mathbf{r}_2)|^2 \tag{1.183}$$

The hopping term is usually limited to nearest neighbors. This Hamiltonian was also introduced by Gutzwiller (1963), who studied the properties of electrons in d-bands in ferromagnets. It was then extensively studied by Hubbard (1963–1966). It is thought to be a good model for electron conduction in narrow band materials, for example, in transition metal

oxides (Adler, 1967). The Hubbard model has been investigated thoroughly over the past forty years, and its properties are starting to be understood. They are discussed in Chapter 6.

The parameter U is the Coulomb interaction between two electrons on the same atom. Usually the model is applied to tightly bound orbitals such as d- or f-states. Then U is quite large, perhaps 6 eV. The bandwidth $\mathcal{Z}w$ is sometimes taken to be smaller. However, some of the most interesting phenomena seem to occur for $U \sim \mathcal{Z}w$. The Hubbard model can be solved exactly only in one dimension, as shown by Lieb and Wu (1968). But there are two limiting cases where exact solutions can be obtained in other dimensions. One is where $U = 0$, which is the nearest neighbor tight-binding model. The other case is where the hopping bandwidth $w = 0$. It is the atomic limit, since here each atom is considered individually, independently of the others. The energy in this atomic limit is

$$E = W_0 N_e + \tilde{n} U \tag{1.184}$$

where N_e is the number of electrons and $\tilde{n}$ is the number of sites with two electrons.

Creation and destruction operators are used to describe other kinds of operators besides Hamiltonians. We shall frequently use density and current operators. The *density operator* is summed over the position of all particles:

$$\rho(\mathbf{r}) = \sum_j \delta(\mathbf{r} - \mathbf{r}_j) \tag{1.185}$$

$\rho(\mathbf{r})$ may be expressed in terms of creation and destruction operators as

$$\rho(\mathbf{r}) = \psi^\dagger(\mathbf{r})\psi(\mathbf{r}) = \sum_{\lambda,\eta} c_\lambda^\dagger c_\eta \phi_\lambda^*(\mathbf{r})\phi_\eta(\mathbf{r}) \tag{1.186}$$

The Fourier transform of the density operator is also needed:

$$\rho(\mathbf{q}) = \int d^3r e^{-i\mathbf{q}\cdot\mathbf{r}} \rho(\mathbf{r}) = \sum_{\lambda,\eta} c_\lambda^\dagger c_\eta \int d^3r \phi_\lambda^*(\mathbf{r})\phi_\eta(\mathbf{r})\, e^{-i\mathbf{q}\cdot\mathbf{r}} \tag{1.187}$$

The two most popular representations are the free-particle model:

$$\rho(\mathbf{q}) = \sum_{\mathbf{k}\sigma} c_{\mathbf{k}+\mathbf{q},\sigma}^\dagger c_{\mathbf{k}\sigma} \tag{1.188}$$

and the tight-binding model when omitting overlap between neighbors:

$$\rho(\mathbf{q}) = \eta(\mathbf{q}) \sum_{j\sigma} n_{j\sigma} e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{1.189}$$

$$\eta(\mathbf{q}) = \int d^3q e^{i\mathbf{q}\cdot\mathbf{r}} |\phi(\mathbf{r})|^2 \tag{1.190}$$

Another important operator is the *electrical current*. It is the summation over all particles and their velocities:

$$\mathbf{j}(\mathbf{r}) = \tfrac{1}{2} \sum_i e_i [\mathbf{v}_i \delta(\mathbf{r} - \mathbf{r}_i) + \delta(\mathbf{r} - \mathbf{r}_i)\mathbf{v}_i] \tag{1.191}$$

The summation above is over different groups of particles. Each group of particles contains identical particles with the same charge e_i. Let $\mathbf{j}_i$ be the *particle current* for each kind of particle species. Then the electrical current operator is

$$\mathbf{j}(\mathbf{r}) = \sum_i e_i \mathbf{j}_i(\mathbf{r}) \tag{1.192}$$

$$\mathbf{j}_i(\mathbf{r}) = \frac{1}{2mi}\{\psi^\dagger(\mathbf{r})\nabla\psi(\mathbf{r}) - \psi(\mathbf{r})\nabla\psi^\dagger(\mathbf{r})\} \tag{1.193}$$

where one sums over each particle species i and the particle current $\mathbf{j}_i$ for each species. The second equation is the standard quantum mechanical representation for the particle current. The Fourier transform of the current operator has the form

$$\mathbf{j}_i(\mathbf{q}) = \frac{1}{2mi}\sum_{\lambda\eta} c_\lambda^\dagger c_\eta \int d^3r e^{-i\mathbf{q}\cdot\mathbf{r}}[\phi_\lambda^*(\mathbf{r})\nabla\phi_\eta(\mathbf{r}) - \phi_\eta(\mathbf{r})\nabla\phi_\lambda^*(\mathbf{r})] \tag{1.194}$$

For free particles, this expression has the form

$$\mathbf{j}_i(\mathbf{q}) = \frac{1}{m}\sum_{\mathbf{k}\sigma}(\mathbf{k} + \tfrac{1}{2}\mathbf{q})c_{\mathbf{k}+\mathbf{q},\sigma}^\dagger c_{\mathbf{k}\sigma} \tag{1.195}$$

Another case of interest is the current operator for the tight binding model. It is easier to consider an alternate, but equivalent, formula for the current operator. The derivation starts from the definition of the *polarization operator* in terms of the particle density $\rho(\mathbf{r})$,

$$\mathbf{P} = \int d^3r\mathbf{r}\rho(\mathbf{r}) \tag{1.196}$$

which is a summation over all the particles and their positions. One then recalls that the time derivative of the polarization is just the particle current:

$$\frac{\partial}{\partial t}\mathbf{P} = \int d^3r\mathbf{r}\frac{\partial}{\partial t}\rho(\mathbf{r}, t) \tag{1.197}$$

This relationship can be proved easily by using the equation of continuity, followed by an integration by parts,

$$\dot{\rho}(\mathbf{r}, t) = -\nabla\cdot\mathbf{j}(\mathbf{r}, t) \tag{1.198}$$

$$\frac{\partial}{\partial t}\mathbf{P} = -\int d^3r\mathbf{r}\nabla\cdot\mathbf{j}(\mathbf{r},\ t) = \int d^3r\mathbf{j}(\mathbf{r})\cdot\nabla\mathbf{r} \tag{1.199}$$

$$= \int d^3r\mathbf{j}(\mathbf{r}, t) \tag{1.200}$$

In the tight-binding model, the polarization operator has the form

$$\mathbf{P} = \sum_i \mathbf{R}_i n_i \tag{1.201}$$

where $\mathbf{R}_i$ is the position of a lattice site, and n_i is the number of particles at that site. The time derivative is

$$\mathbf{j} = \frac{\partial}{\partial t}\mathbf{P} = i[H, \mathbf{P}] \tag{1.202}$$

For example, if the Hamiltonian has the form

$$H = w\sum_{j\delta\sigma} c^{\dagger}_{j+\delta,\sigma}c_{j\sigma} + \frac{1}{2}\sum_{ij,ss'} n_{is}n_{js'}V_{ij} \tag{1.203}$$

then only the first term contributes to the current operator. The other term contains only the position operator n_{is}, which commutes with itself. Here s is the spin index. In this case the current operator is

$$\mathbf{j} = -iw\sum_{j\delta\sigma} \delta c^{\dagger}_{j+\delta,\sigma}c_{j\sigma} \tag{1.204}$$

This current operator is used in calculations on localized electrons. It applies to organic solids and narrow band ionic solids.

The *energy current* operator is needed to calculate energy transport in solids, which occurs, for example, in discussions of thermal conductivity or thermoelectric effects. Energy currents flow whenever heat is generated or dissipated nonuniformly in the solid. The energy current $\mathbf{j}_E$ is defined as the energy flow through a surface. It obeys an equation of energy conservation,

$$\frac{\partial}{\partial t}H + \nabla\cdot\mathbf{j}_E = 0 \tag{1.205}$$

where the energy change $\partial H/\partial t$ equals the variation in the energy flux. An equation for the energy current may be derived by formally introducing an operator which is the integral over the position and Hamiltonian density:

$$\mathbf{R}_E = \tfrac{1}{2}\int d^3r[\mathbf{r}\mathcal{H}(\mathbf{r}) + \mathcal{H}(\mathbf{r})\mathbf{r}] \tag{1.206}$$

The equation of energy continuity (1.205) may be used to show that the time derivative of this quantity is just the energy current:

$$\frac{\partial}{\partial t}\mathbf{R}_E = \frac{1}{2}\int d^3r\left[\mathbf{r}\frac{\partial}{\partial t}\mathcal{H}(\mathbf{r}) + \frac{\partial}{\partial t}\mathcal{H}(\mathbf{r})\mathbf{r}\right] = \mathbf{j}_E \tag{1.207}$$

For a free-particle system, the energy current is

$$\mathbf{j}_E = \sum_{\mathbf{p}\sigma} \mathbf{v}_{\mathbf{p}}\varepsilon_{\mathbf{p}}c^{\dagger}_{\mathbf{p}\sigma}c_{\mathbf{p}\sigma} \tag{1.208}$$

This result is sensible. It is just the energy $\varepsilon_{\mathbf{p}} = p^2/2m$ of each particle multiplied by its velocity $\mathbf{v}_{\mathbf{p}} = \mathbf{p}/m$. In the nearest neighbor tight-binding model, one writes $\mathbf{R}_E$ in terms of the site Hamiltonian h_i and position $\mathbf{R}_i$:

$$\mathbf{R}_E = \sum_i \mathbf{R}_i h_i \tag{1.209}$$

$$H = \sum_i h_i \tag{1.210}$$

$$\mathbf{j}_E = i\sum_{lm} \mathbf{R}_l[h_m, h_l] \tag{1.211}$$

$$h_i = \frac{w}{2}\sum_{\delta\sigma}(c^{\dagger}_{i+\delta,\sigma}c_{i\sigma} + c^{\dagger}_{i\sigma}c_{i+\delta,\sigma}) + \frac{1}{2}\sum_{jss'} n_{is}n_{js'}V_{ij} \tag{1.212}$$

For example, the current operator from just the hopping term for electrons is in one dimension:

$$\mathbf{j}_E = -i\frac{w^2}{2}\sum_{i\delta\delta' s}(\delta+\delta')\, c^{\dagger}_{i+\delta+\delta',s}c_{is} \tag{1.213}$$

The other terms can become quite complicated, and will be introduced only as they are needed.

There are several other comments. First, the operators $\mathbf{P}$ and $\mathbf{R}_E$ are not defined in infinite systems, since the integral over position will diverge. One can devise an alternate definition

$$\mathbf{P} = -i\lim_{\mathbf{q}\to 0}\nabla_{\mathbf{q}}\int d^3r\rho(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}} \tag{1.214}$$

$$\mathbf{R}_E = -i\lim_{\mathbf{q}\to 0}\nabla_{\mathbf{q}}\int d^3r\{e^{-i\mathbf{q}\cdot\mathbf{r}}, \mathcal{H}(\mathbf{r})\} \tag{1.215}$$

The second comment concerns the energy current. As discussed extensively in Sec. 3.8, the energy current is often not the current which describes thermal conductivity or thermoelectric power. In metals, it is customary to use the *heat current*, which is defined as

$$\mathbf{j}_Q = \mathbf{j}_E - \mu\mathbf{j} \tag{1.216}$$

is given in terms of operators $\mathbf{j}_E$ and $\mathbf{j}$, which have already been defined.

Another observation is that these operator definitions are the same for all particles, regardless of whether they are bosons or fermions. Of course, spinless bosons do not have the summation over spin index σ. Otherwise, everything is the same. However, calculations using these operators depend significantly on whether the particles are bosons or fermions.

1.3. ELECTRON–PHONON INTERACTIONS

The first two sections described the Hamiltonians for phonons and electrons, respectively. This section discusses their mutual interaction. This topic is important in many-body theory. The electron–phonon interaction causes superconductivity in many metals and influences the transport properties of every metal. In pure semiconducting and ionic solids, the electron–phonon interaction usually dominates the transport properties. The word *polaron* is used to describe a single electron which is coupled to phonons. The modern formulation of the polaron problem was due to Fröhlich *et al.* (1950), and its study is an important part of the history of many-body theory.

1.3.1. Interaction Hamiltonian

The basic Hamiltonian is assumed to have the form

$$H = H_p + H_e + H_{ei} \tag{1.217}$$

$$H_p = \sum_{\mathbf{q}\lambda} \omega_{\mathbf{q}\lambda} a^{\dagger}_{\mathbf{q}\lambda} a_{\mathbf{q}\lambda} \tag{1.218}$$

$$H_e = \sum_i \left[\frac{p_i^2}{2m} + \frac{e^2}{2} \sum_{j \neq i} \frac{1}{r_{ij}} \right] \tag{1.219}$$

$$H_{ei} = \sum_{ij} V(\mathbf{r}_i - \mathbf{R}_j) \tag{1.220}$$

The atom part H_p describes the normal modes of vibration of the solid and is the phonon Hamiltonian of Sec. 1.1. The second term is the electron part H_e, which was discussed in Sec. 1.2. The third part is the electron–ion interaction. It is assumed that H_{ei} is the summation of the interaction between the individual atoms at $\mathbf{R}_j$ and the electrons at $\mathbf{r}_i$: The word ion is not meant to imply a particular charge state. In metals, the atoms are ions, while in covalently bonded semiconductors they are something else. The word *ion* is used to encompass all these possibilities. Each ion is at a position $\mathbf{R}_j = \mathbf{R}_j^{(0)} + \mathbf{Q}_j$, which is the sum of the equilibrium position $\mathbf{R}_j^{(0)}$ and the displacement $\mathbf{Q}_j$. The displacements are usually small, so that one can expand in powers of them:

$$V_{ei}(\mathbf{r}_i - \mathbf{R}_j^{(0)} - \mathbf{Q}_j) = V_{ei}(\mathbf{r}_i - \mathbf{R}_j^{(0)}) - \mathbf{Q}_j \cdot \nabla V_{ei}(\mathbf{r}_i - \mathbf{R}_j^{(0)}) + O(Q^2)$$

The linear electron–phonon interaction term is obtained from the first term in $\mathbf{Q}_j$. The terms in $O(Q^2)$ are neglected here, although they are retained in some circumstances. The constant term

$$\sum_j V_{ei}(\mathbf{r}_i - \mathbf{R}_j^{(0)}) \tag{1.221}$$

is the potential function for the electrons when the atoms are in their equilibrium positions, which forms a periodic potential in a crystal. The solution of the Hamiltonian for electron motion in this periodic potential gives the Bloch states of the solids. They are usually assumed to be known. Many problems begin by writing the Hamiltonian as $H = H_0 + V$, where H_0 is a Hamiltonian which can be solved and V is the perturbation. Quite often the eigenstates of H_0 are just the Bloch states of the solid, calculated by assuming that the atoms are in their equilibrium positions. The sequence of approximations we are making is called, collectively, the *Born–Oppenheimer approximation*. As always, there are circumstances where these approximations are inadequate, and other approaches are necessary, e.g., in the dynamical Jahn–Teller effect.

The electron–phonon interaction is

$$V_{ep}(\mathbf{r}) = \sum_j \mathbf{Q}_j \cdot \nabla V_{ei}(\mathbf{r} - \mathbf{R}_j^{(0)}) \tag{1.222}$$

This interaction is to be written in terms of operators. It is assumed that the electron–atom potential possesses a Fourier transform:

$$V_{ei}(\mathbf{r}) = \frac{1}{N}\sum_q V_{ei}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.223}$$

$$\nabla V_{ei}(\mathbf{r}) = \frac{i}{N}\sum_q \mathbf{q}V_{ei}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.224}$$

Evaluate the combination

$$V_{ep}(\mathbf{r}) = \frac{i}{N}\sum_q V_{ei}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{r}}\mathbf{q}\cdot\left(\sum_j \mathbf{Q}_j e^{-i\mathbf{q}\cdot\mathbf{R}_j^{(0)}}\right) \tag{1.225}$$

The earlier definition (1.85) of $\mathbf{Q}_j$ is used to show that

$$\frac{i}{N}\sum_j \mathbf{Q}_j e^{-i\mathbf{q}\cdot\mathbf{R}_j^{(0)}} = \frac{i}{\sqrt{N}}\sum_{\mathbf{G}} \mathbf{Q}_{\mathbf{q}+\mathbf{G}} \tag{1.226}$$

$$= -\sum_{\mathbf{G}}\left(\frac{\hbar}{2MN\omega_{\mathbf{q}}}\right)^{1/2}\xi_{\mathbf{q}+\mathbf{G}}(a_{\mathbf{q}} + a_{\mathbf{q}}^{\dagger}) \tag{1.227}$$

where the summation $\mathbf{G}$ is over the reciprocal lattice vectors of the solid. The phonon states of $\mathbf{q}+\mathbf{G}$ are defined only within the first Brillouin zone of the solid. Here the values of $\mathbf{q}$ may be outside the zone, so that $\mathbf{q}+\mathbf{G}$ projects them back into it. But the Fourier transform $V_{ei}(\mathbf{q}+\mathbf{G})$ is defined over all values of $(\mathbf{q}+\mathbf{G})$, not just the first Brillouin zone. Write the interaction Hamiltonian in the form ($MN = \rho v$, ρ = density of solid in grams per cubic centimeter)

$$V_{ep}(\mathbf{r}) = -\sum_{\mathbf{qG}} e^{i\mathbf{r}\cdot(\mathbf{q}+\mathbf{G})}V_{ei}(\mathbf{q}+\mathbf{G})(\mathbf{q}+\mathbf{G})\cdot\xi_{\mathbf{q}}\left(\frac{\hbar}{2\rho v\omega_{\mathbf{q}}}\right)^{1/2}(a_{\mathbf{q}} + a_{-\mathbf{q}}^{\dagger})$$

The summation over $\mathbf{q}$ is restricted to be within the first Brillouin zone of the crystal. The phonons are defined only in this space, so $\omega_{\mathbf{q}}$, $a_{\mathbf{q}}$, $\xi_{\mathbf{q}}$ have only $\mathbf{q}$ labels. But the summation over reciprocal lattice vectors $\mathbf{G}$ permits the potential $V_{ei}(\mathbf{q}+\mathbf{G})$ to interact with higher Fourier components. The notation also dropped the summation over the various phonon modes λ. Such a summation should be done, and we always do it, but writing λ in every subscript is a nuisance. So just do the summation over modes, and skip the writing of λ in every subscript!

The potential $V_{ep}(\mathbf{r})$ is defined as the unscreened electron–atom potential. Later it is shown that electron–electron interactions in metals cause a significant reduction of this potential, which is called screening. The potential V_{ep} is sometimes calculated from first principles, but more often it is obtained from a pseudopotential (Heine, 1970; Harrison, 1966). Modern calculations usually use force constant models to obtain accurate phonon energies $\omega_{\mathbf{q}}$, and polarizations $\xi_{\mathbf{q}}$ throughout the Brillouin zone, for each mode of polarization and use them in calculating electron–phonon properties.

The potential $V_{ep}(\mathbf{r})$ acts upon the electrons and also upon other particles such as positrons. The electron–phonon interaction is obtained by integrating this potential over the charge density of the solid $\rho(\mathbf{r})$:

$$H_{ep} = \int d^3r\rho(\mathbf{r})V_{ep}(\mathbf{r})$$

$$= -\sum_{\mathbf{qG}} \rho(\mathbf{q}+\mathbf{G})V_{ei}(\mathbf{q}+\mathbf{G})(\mathbf{q}+\mathbf{G})\cdot\xi_{\mathbf{q}}\left(\frac{\hbar}{2\rho v\omega_{\mathbf{q}}}\right)^{1/2}(a_{\mathbf{q}}+a^{\dagger}_{-\mathbf{q}})$$

The particle density operator $\rho(\mathbf{q})$ was defined in the prior section. Several examples of it were given there. Quite often this matrix element is abbreviated by the symbol

$$M_{\mathbf{q}+\mathbf{G}} = -V_{ei}(\mathbf{q}+\mathbf{G})(\mathbf{q}+\mathbf{G})\cdot\xi_{\mathbf{q}}\left(\frac{\hbar}{2\rho v\omega_{\mathbf{q}}}\right)^{1/2} \tag{1.228}$$

$$H_{ep} = \sum_{\mathbf{qG}} M_{\mathbf{q}+\mathbf{G}}\rho(\mathbf{q}+\mathbf{G})(a_{\mathbf{q}}+a^{\dagger}_{-\mathbf{q}}) \tag{1.229}$$

1.3.2. Localized Electron

There is one problem which can be solved immediately: the electron–phonon Hamiltonian when the electrons are fixed in space at positions $\mathbf{r}_i$. It is assumed that they cannot recoil, which neglects the electron kinetic energy term. This model is often applied for localized electrons in solids. A localized electron occurs in deep core states and in some impurity levels. Then the Hamiltonian has the form

$$H = H_p + H_{ep} \tag{1.230}$$

$$= \sum_{\mathbf{q}}\left[\omega_{\mathbf{q}}\left(a^{\dagger}_{\mathbf{q}}a_{\mathbf{q}}+\frac{1}{2}\right)+\sum_i(a_{\mathbf{q}}+a^{\dagger}_{-\mathbf{q}})e^{i\mathbf{q}\cdot\mathbf{r}_i}\right.$$

$$\left.\times\sum_{\mathbf{G}} M_{(\mathbf{q}+\mathbf{G})}\eta(\mathbf{q}+\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{r}_i}\right] \tag{1.231}$$

The electron density operator $\eta(\mathbf{q}+\mathbf{G})$ is just the Fourier transform of the localized charge density:

$$\rho(\mathbf{q}+\mathbf{G}) = \int d^3re^{i\mathbf{r}\cdot(\mathbf{q}+\mathbf{G})}\sum_i|\phi_0(\mathbf{r}-\mathbf{r}_i)|^2 \tag{1.232}$$

$$= \sum_i e^{i\mathbf{r}_i\cdot(\mathbf{q}+\mathbf{G})}\eta(\mathbf{q}+\mathbf{G}) \tag{1.233}$$

$$\eta(\mathbf{q}+\mathbf{G}) = \int d^3re^{i\mathbf{r}\cdot(\mathbf{q}+\mathbf{G})}|\phi_0(\mathbf{r})|^2 \tag{1.234}$$

The various interaction terms are collected into an effective matrix element:

$$F_{\mathbf{q}}(\mathbf{r}) = \sum_{\mathbf{G}} \eta(\mathbf{q}+\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} M_{(\mathbf{q}+\mathbf{G})} \tag{1.235}$$

$$H = \sum_{\mathbf{q}} \left[\omega_{\mathbf{q}} \left(a_{\mathbf{q}}^{\dagger} a_{\mathbf{q}} + \frac{1}{2} \right) + \sum_{i} (a_{\mathbf{q}} + a_{-\mathbf{q}}^{\dagger}) e^{i\mathbf{q}\cdot\mathbf{r}_i} F_{\mathbf{q}}(\mathbf{r}_i) \right] \tag{1.236}$$

The function $F_{\mathbf{q}}(\mathbf{r})$ is periodic in the lattice, since increasing it by a lattice vector **a** does not change its value. It is assumed that all localized electrons in different unit cells are in the same position within the cell. That is, $F_{\mathbf{q}}(\mathbf{r})$ is assumed to be the same for all localized electrons.

This problem has been solved before. It is just the harmonic oscillator in an electric field [Eq. (1.38)]. Now each wave vector and polarization state **q** is a separate harmonic oscillator, which finds its own equilibrium configuration. Follow exactly the steps used to solve (1.38), and the creation and destruction operators are transformed to the new set

$$A_{\mathbf{q}} = a_{\mathbf{q}} + \frac{F_{\mathbf{q}}}{\omega_{\mathbf{q}}} \sum_{i} e^{i\mathbf{q}\cdot\mathbf{r}_i} \tag{1.237}$$

$$A_{\mathbf{q}}^{\dagger} = a_{\mathbf{q}} + \frac{F_{\mathbf{q}}^{*}}{\omega_{\mathbf{q}}} \sum_{i} e^{-i\mathbf{q}\cdot\mathbf{r}_i} \tag{1.238}$$

The Hamiltonian with these operators is

$$H = \sum_{\mathbf{q}} \omega_{\mathbf{q}} \left(A_{\mathbf{q}}^{\dagger} A_{\mathbf{q}} + \frac{1}{2} \right) - \sum_{\mathbf{q}} \frac{|F_{\mathbf{q}}|^2}{\omega_{\mathbf{q}}} |\sum_{i} e^{i\mathbf{q}\cdot\mathbf{r}_i}|^2 \tag{1.239}$$

Furthermore, the new operators still obey the harmonic oscillator commutation relations

$$[A_{\mathbf{q}}, A_{\mathbf{q}'}^{\dagger}] = \delta_{\mathbf{q}\mathbf{q}'} \tag{1.240}$$

$$[A_{\mathbf{q}}, A_{\mathbf{q}'}] = 0 \tag{1.241}$$

$$[A_{\mathbf{q}}^{\dagger}, A_{\mathbf{q}'}^{\dagger}] = 0 \tag{1.242}$$

It was shown in Sec. 1.1 that the eigenstates and eigenvalues for this Hamiltonian are

$$|n_{\mathbf{q}}\rangle = \frac{(A_{\mathbf{q}}^{\dagger})^{n_{\mathbf{q}}}}{\sqrt{n_{\mathbf{q}}!}} |0\rangle \tag{1.243}$$

$$E = \sum_{\mathbf{q}} \omega_{\mathbf{q}} \left(n_{\mathbf{q}} + \frac{1}{2} \right) - \sum_{\mathbf{q}} \frac{|F_{\mathbf{q}}|^2}{\omega_{\mathbf{q}}} |\sum_{i} e^{i\mathbf{q}\cdot\mathbf{r}_i}|^2 \tag{1.244}$$

The eigenstates are interpreted in the same way as used for the simple spring: Each normal mode **q** has stretched to a new equilibrium configuration,

$$\mathbf{Q}_{\mathbf{q}}^{(0)} = -2 \left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}} \right)^{1/2} \frac{F_{\mathbf{q}}}{\omega_{\mathbf{q}}} \sum_{i} e^{i\mathbf{q}\cdot\mathbf{r}_i} \tag{1.245}$$

and now oscillates about this new equilibrium point. The oscillation frequencies do not change. The last term in (1.244) is the *relaxation energy*. It is the potential energy gained by

stretching the springs—the phonon normal modes—to the new equilibrium positions. This energy term may be expanded in terms of the electron coordinates. The relaxation energy is

$$= -\sum_{\mathbf{q}} \frac{|F_{\mathbf{q}}|^2}{\omega_{\mathbf{q}}} \sum_{ij} e^{i\mathbf{q}\cdot(\mathbf{r}_i - \mathbf{r}_j)} = \frac{1}{2} \sum_{ij} V_R(\mathbf{r}_i - \mathbf{r}_j) \tag{1.246}$$

$$= \frac{1}{2} \sum_{i} V_R(0) + \sum_{i>j} V_R(\mathbf{r}_i - \mathbf{r}_j) \tag{1.247}$$

$$V_R(\mathbf{r}) = -2 \sum_{\mathbf{q}} \frac{|F_{\mathbf{q}}|^2}{\omega_{\mathbf{q}}} e^{i\mathbf{q}\cdot\mathbf{r}} = -2v \sum_{\lambda} \int \frac{d^3q}{(2\pi)^3} \frac{|F_{\mathbf{q}\lambda}|^2}{\omega_{\mathbf{q}\lambda}} e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.248}$$

Note that the summation over phonon polarizations has been added to the last expression, which is the integral one has to evaluate.

The relaxation energy consists of two types of terms. The first is $V_R(0)$, which is the relaxation energy of a single particle by itself. This energy is caused by the electron inducing a static polarization in the phonon field which acts back upon the electron. It is a self-energy effect. It shall be called a polaron self-energy or electron self-energy. The energy is not just with the electron. As with the stretched spring, it is in the combined particle–oscillator system.

The other type of term is the interaction between pairs of fixed particles $V_R(\mathbf{r}_i - \mathbf{r}_j)$. Here the physical picture is that one particle polarizes the medium, and this polarization field changes the energy of other particles which are nearby. This potential $V_R(\mathbf{r})$ has different $\mathbf{r}$ dependence for different types of phonons in solids. In some cases, it is very short-ranged, so that two particles interact only when they are in the same unit cell of the crystal. In other cases, the potential falls as slowly such as r^{-1}, as if it were a Coulomb potential. Several of these cases will be presented later.

The same many-body problem is discussed again in Chapter 4. The harmonic oscillator in a linear potential is an important model, if only because there are so few models which can be solved exactly. It is also the solution to the independent boson model.

1.3.3. Deformation Potential

In semiconductors and ionic solids, the excited electron states are usually confined to a small location in wave vector space. In thermal equilibrium, the excited states are at an energy band minimum, which is often at the zone center or edge. Usually the polaron effects for these electrons involve only phonons of long wavelength. In this case the tradition has been to parameterize the interaction rather than compute it from first principles. Most electron–phonon interactions in semiconductors use only three types of interactions: deformation potential coupling to acoustical phonons, piezoelectric coupling to acoustical phonons, and polar coupling to optical phonons. Another possible coupling is the deformation coupling to optical phonons. These interactions are valid only at long wavelength. When the electron–phonon matrix element is needed at short-wavelength phonons, the usual method is to calculate them from pseudopotentials.

The deformation potential coupling to acoustical phonons is just the long-wavelength limit of (1.229). Only the $\mathbf{G} = 0$ is retained, since the terms $\mathbf{G} \neq 0$ are of wavelengths that are too short. The electron–ion potential is taken to be a constant ($V_{ei}(\mathbf{q}) \to D$) at $\mathbf{q} \to 0$, where

D is the *deformation constant*. At long wavelength, $\xi \to \hat{\mathbf{q}}$, and only longitudinal phonons are important if the band is nondegenerate. The interaction has the form

$$H_{ep} = D\sum_{\mathbf{q}}\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}\mathrm{v}}\right)^{1/2} |\mathbf{q}|\rho(\mathbf{q})(a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}}) \tag{1.249}$$

where the summation is only over longitudinal acoustic phonons. However, valence bands in semiconductors are often degenerate at the band maximum. Then the hole excitations have a deformation coupling to transverse phonons, which is a very large polaron correction (Mahan, 1965).

The deformation constants are obtained by measuring how energy bands shift with increasing pressure on the solid. The value of D_n for a band n is simply the rate of change of band energy with pressure (Thomas, 1961).

1.3.4. Piezoelectric Interaction

Many semiconductors are piezoelectric. The macroscopic effect is that an electric field is generated when a crystal is squeezed and vice versa. Acoustical phonons, which are periodic density modulations, make periodic electric fields. The crystal must lack an inversion center to be piezoelectric. The group IV semiconductors Ge and Si are not piezoelectric. The III–V semiconductors such as GaAs are very weakly piezoelectric, while the II–VI materials such as CdS and ZnO are extremely piezoelectric. A very detailed derivation of the electron–phonon interaction has been given elsewhere (Mahan, 1972). Here we shall provide only a quick sketch. If S_{ij} is the stress on the crystal, then the electric field is proportional to the stress,

$$E_k = \sum_{ij} M_{ijk} S_{ij} \tag{1.250}$$

where the matrix M_{ijk} is a constant which gives the proportionality. The stress is defined as the symmetric derivative of the displacement field:

$$S_{ij} = \frac{1}{2}\left(\frac{\partial Q_i}{\partial x_j} + \frac{\partial Q_j}{\partial x_i}\right) \tag{1.251}$$

$$= \frac{1}{2}\sum_{\mathbf{q}}\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}\mathrm{v}}\right)^{1/2} (\xi_i q_j + \xi_j q_i)(a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}})e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.252}$$

The electric field may be shown to be longitudinal and to point in the direction $\mathbf{q}$ of the phonon. It may be written as the gradient of a potential $\phi(\mathbf{r})$

$$E_k = -\frac{\partial}{\partial x_k}\phi(\mathbf{r}) = -\frac{1}{\sqrt{\mathrm{v}}}\sum_{\mathbf{q}} iq_k\phi_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.253}$$

The final result is the observation that the potential is proportional to the displacement:

$$\phi(\mathbf{r}) \propto Q(\mathbf{r}) \tag{1.254}$$

$$\phi(\mathbf{r}) = i\sum_{\mathbf{q}}\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}\mathrm{v}}\right)^{1/2} M_{\lambda}(\hat{q})(a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}})e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.255}$$

The electron–phonon interaction for the piezoelectric interaction is

$$H_{ep} = i\sum_{\mathbf{q}}\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}\mathrm{v}}\right)^{1/2} M_\lambda(\hat{q})\rho(\mathbf{q})(a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}}) \tag{1.256}$$

The matrix element $M_\lambda(\hat{q})$ does not depend on the magnitude of $\mathbf{q}$, but it very much depends on its direction. It also has the property that $M_\lambda(-\hat{q}) = -M_\lambda(\hat{q})$ so that (1.249) is Hermitian. In fact, the piezoelectric interaction is quite anisotropic. The matrix element is also very dependent on the polarization λ of the acoustical phonon whether it is LA (longitudinal acoustic) or TA (transverse acoustic). Most many-body calculations have tended to take a constant value for the matrix element, where this constant is obtained by averaging over the various angular directions in the crystal. This approximation is adopted here.

An interesting result is obtained if (1.256) is used to calculate the effective potential energy between two fixed electrons:

$$V_R(\mathbf{r}) = -2\int\frac{d^3q}{(2\pi)^3}\sum_\lambda\frac{M_\lambda^2}{2\rho\omega_{\mathbf{q}\lambda}^2} \tag{1.257}$$

In a Debye model, the phonon energy is proportional to wave vector $\omega_{\mathbf{q}\lambda} = c_\lambda q$, where the sound velocity c_λ depends upon the mode. Since M_λ and ρ are constants, the potential $V_R(r)$ is the Fourier transform of q^{-2}. This transform was worked out earlier, in (1.163), and just gives a potential varying as r^{-1}:

$$V_R(r) = -\gamma\frac{e^2}{r}$$

$$\gamma = \frac{1}{4\pi e^2\rho}\sum_\lambda\frac{M_\lambda^2}{c_\lambda^2} \tag{1.258}$$

The answer is a form of Coulomb's law. Actually, it is a law of dielectric screening. In dielectric materials it is known that the potential between two fixed charges is

$$V_{\text{total}}(r) = \frac{e^2}{r\varepsilon_{\text{total}}} \tag{1.259}$$

where $\varepsilon_{\text{total}}$ is the total, static, dielectric constant. In the piezoelectric calculation, the piezoelectric contribution to this dielectric screening was derived. The total dielectric function is the summation of many contributions:

$$\varepsilon_{\text{total}} = \varepsilon_\infty + \varepsilon_{\text{piezo}} + \varepsilon_{\text{polar}} + \varepsilon_{e-e} \tag{1.260}$$

$$\varepsilon_{\text{total}} = \varepsilon_0 + \varepsilon_{\text{piezo}} \tag{1.261}$$

The first term ε_∞ is from interband electronic transitions. The piezoelectric contribution is the term under discussion. The polar and electron–electron contributions are explained later. All other contributions are ε_0. All of these contributions are functions of $\mathbf{q}$ and ω. In the present problem, with a static, fixed charge, these quantities are needed in the limit that $\mathbf{q} \to 0$ and $\omega \to 0$. The electron–electron term is not finite in this limit, so probably should not have been included in the list (1.260). It may be omitted if there are no mobile charges present, which shall be assumed.

If there were no piezoelectric contribution, presumably the interaction potential between two fixed charges would be $e^2/(\varepsilon_0 r)$. The piezoelectric contribution (1.258) represents the difference between the above result and total screening (1.259):

$$\frac{e^2}{\varepsilon_{\text{total}} r} = \frac{e^2}{r}\left(\frac{1}{\varepsilon_0} - \gamma\right) \tag{1.262}$$

so that

$$\frac{1}{\varepsilon_0 + \varepsilon_{\text{piezo}}} = \frac{1}{\varepsilon_0} - \gamma \tag{1.263}$$

These equations can be solved to obtain the piezoelectric contribution to the static dielectric function:

$$\varepsilon_{\text{piezo}} = \frac{\varepsilon_0^2 \gamma}{1 - \varepsilon_0 \gamma} \tag{1.264}$$

The interaction term (1.258) is negative because the screening lowers the potential energy. The unscreened potential is just e^2/r, and each bit of screening lowers it by an amount proportional to e^2/r.

The same acoustical phonon may interact with an electron by both the deformation and piezoelectric interactions. These two interactions do not interfere, to second order, because they are out of phase. The sum of the two interactions (1.249) and (1.256) gives

$$H_{ep} = \sum_{\mathbf{q}} \left(\frac{\hbar}{2\rho\omega_{\mathbf{q}} V}\right)^{1/2} \bar{M}(\mathbf{q})\rho(\mathbf{q})(a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}}) \tag{1.265}$$

$$\bar{M}(\mathbf{q}) = D|\mathbf{q}| + iM_{\lambda}(\hat{q}) \tag{1.266}$$

The deformation potential is real, while the piezoelectric is imaginary. To second order, they do not interfere:

$$|\bar{M}|^2 = D^2 q^2 + M_{\lambda}^2 \tag{1.267}$$

Since electron–phonon effects are usually evaluated only to second order, these two interactions can be treated separately.

1.3.5. Polar Coupling

The polar coupling between electrons and optical phonons can be very large in ionic crystals. The form of the Hamiltonian has been derived often, e.g., Fröhlich (1954). In ionic crystals some of the atoms are positively charged, while others are negatively charged. An optical phonon has the different ions in the crystal vibrating out of phase. When the plus ions and minus ions oscillate in the opposite direction, they set up a dipole polarization field. The polarization causes a long-range electric field which scatters the electrons. The electric field is the source of the polar coupling.

The polar coupling is only to LO (longitudinal optical) phonons and not to TO (transverse optical) phonons (Mahan, 1972), because only the LO phonons set up strong electric

fields when they vibrate. These electric fields are in the direction of vibration, which at long wavelength is in the direction of the phonon wave vector $\mathbf{q}$. For a system of no free charges,

$$\nabla \cdot \mathbf{D} = 0 = \sum_{\mathbf{q}} \mathbf{q} \cdot (\mathbf{E}_{\mathbf{q}} + 4\pi \mathbf{P}_{\mathbf{q}}) e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.268}$$

For an LO phonon mode of wave vector $\mathbf{q}$, the electric field $\mathbf{E}_{\mathbf{q}}$ and polarization $\mathbf{P}_{\mathbf{q}}$ are both parallel to $\mathbf{q}$. The formula for the electric field produced by the polarization is:

$$\mathbf{E}_{\mathbf{q}} = -4\pi \mathbf{P}_{\mathbf{q}} \tag{1.269}$$

The next assumption is that the polarization is proportional to the displacement

$$\mathbf{P}_{\mathbf{q}} = eU\mathbf{Q}_{\mathbf{q}} \tag{1.270}$$

$$\mathbf{E}_{\mathbf{q}} = -4\pi eU\mathbf{Q}_{\mathbf{q}} = -4\pi ieU\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}\nu}\right)^{1/2} \hat{q}(a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}}) \tag{1.271}$$

where the coefficient U is to be determined. The phonon energy ω_{LO} is assumed to be constant. Since the electric field points in the direction of $\mathbf{q}$, it may be expressed by a potential:

$$\mathbf{E} = -\nabla\phi = -i\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \mathbf{q}\phi_{\mathbf{q}} \tag{1.272}$$

$$\phi(\mathbf{r}) = \sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \frac{4\pi eU}{q}\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}\nu}\right)^{1/2} \hat{q}(a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}}) \tag{1.273}$$

which gives the potential produced by the LO phonons. The interaction constant U still needs to be determined. Its value is obtained by considering the potential between two fixed electrons that was calculated from our relaxation energy (1.247):

$$V_R(r) = -\frac{2}{\hbar\omega_{LO}}(4\pi eU)^2\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}}\right)\int \frac{d^3q}{(2\pi)^3}\frac{e^{i\mathbf{q}\cdot\mathbf{r}}}{q^2} \tag{1.274}$$

The Fourier transform integral again produces a Coulomb potential:

$$V_R(r) = -\Gamma\frac{e^2}{r} \tag{1.275}$$

$$\Gamma = \frac{4\pi U^2}{\rho\omega_{LO}^2} \tag{1.276}$$

This interaction energy is again interpreted as a contribution to the dielectric screening of the solid. This term represents the contribution from the optical phonons. It represents the difference between screening with just the electronic interband part ε_∞ and the interband plus optical phonons ε_0:

$$\frac{e^2}{r\varepsilon_0} = \frac{e^2}{r}\left(\frac{1}{\varepsilon_\infty} - \Gamma\right) \tag{1.277}$$

$$\Gamma = \frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0} \tag{1.278}$$

The unknown factor U is

$$U^2 = \frac{\rho\omega_{LO}^2}{4\pi}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \tag{1.279}$$

It is possible to write the electron–phonon Hamiltonian in the form

$$H_{ep} = \sum_{\mathbf{q}} \frac{M}{q\sqrt{\nu}}\rho(\mathbf{q})(a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}}) \tag{1.280}$$

$$M^2 = 2\pi e^2\hbar\omega_{LO}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \tag{1.281}$$

where the matrix element M is given in the second line as M^2. This form is a very sensible way in which to express the matrix element. The dielectric constants ε_0 and ε_∞ are both measurable: ε_0 is the low-frequency dielectric function measured by putting the solid between the parallel plates of a capacitor at low frequency, while ε_∞ is the square of the refractive index.

Besides these quantities, the matrix element depends only on the charge e and the LO phonon energy $\hbar\omega_{LO}$. In spite of the elegance of this simple form, the matrix element is not usually expressed this way. Instead, it is customary to introduce the dimensionless polaron constant α, defined as

$$\alpha = \frac{e^2}{\hbar}\left(\frac{m}{2\hbar\omega_{LO}}\right)^{1/2}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \tag{1.282}$$

Correspondingly, the interaction matrix element is

$$M^2 = \frac{4\pi\alpha\hbar(\hbar\omega_{LO})^{3/2}}{\sqrt{2m}} \tag{1.283}$$

The reason for introducing α is simple. The self-energy expressions will have a simple and elegant form with this choice of α.

Electrons in crystalline energy bands have their motion determined by an effective band mass m_b. The band mass enters into the definition of α, which has the obvious disadvantage that one does not know the value of α until one knows the value of m_b.

1.4. SPIN HAMILTONIANS

The study of spin systems forms a very large part of many-body theory. There are many solids which display magnetic ordering among the electrons. There are many impurity problems where spin plays an important role. In attempting to explain these phenomena, many different types of spin models have been introduced. Some of these involve localized spins interacting among themselves, while others have localized spins interacting with free electrons. It is a subject in which there are few exactly solvable models which are nontrivial. Only a few of the models have solutions which are well understood, in spite of the fact that many of them have been intensely studied. The transference of model results to real solids which are strongly interacting systems has still not been very successful.

The commutation relations for spin one-half operators are ($\hbar = 1$)

$$[S_l^{(x)}, S_j^{(y)}] = iS_j^{(z)}\delta_{lj}$$

$$[S_l^{(y)}, S_j^{(z)}] = iS_j^{(x)}\delta_{lj}$$

$$[S_l^{(z)}, S_j^{(x)}] = iS_j^{(y)}\delta_{lj} \tag{1.284}$$

The subscript label l or j denotes spin site. Spins on different sites, or with different electrons, commute. The superscripts (x), (y), (z) refer to space coordinates. These spin operators are often represented by Pauli spin matrices,

$$S^{(x)} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$S^{(y)} = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$S^{(z)} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.285}$$

but these representations are unnecessary for many-body calculations—the commutation relations are sufficient. It is customary to introduce spin raising and lowering operators:

$$S_j^{(+)} = S_j^{(x)} + iS_j^{(y)} \tag{1.286}$$

$$S_j^{(-)} = S_j^{(x)} - iS_j^{(y)} \tag{1.287}$$

The names raising and lowering are applied to these operators because they raise or lower the magnetic quantum number m of the spin state. For spin one-half, there is only spin up $|+\rangle$ and spin down $|-\rangle$, and the operators go from one to the other:

$$S^{(+)}|-\rangle = |+\rangle, \quad S^{(-)}|+\rangle = |-\rangle$$

$$S^{(+)}|+\rangle = 0, \quad S^{(-)}|-\rangle = 0 \tag{1.288}$$

By direct multiplication, for two operators on the same site:

$$S^{(+)}S^{(-)} = S^{(x)2} + S^{(y)2} - i[S^{(x)}S^{(y)} - S^{(y)}S^{(x)}] = S^{(x)2} + S^{(y)2} + S^{(z)}$$

$$S^{(-)}S^{(+)} = S^{(x)2} + S^{(y)2} + i[S^{(x)}S^{(y)} - S^{(y)}S^{(x)}] = S^{(x)2} + S^{(y)2} - S^{(z)} \tag{1.289}$$

By subtracting these two results, the commutation relations are found among the raising and lowering operators:

$$[S_l^{(+)}, S_j^{(-)}] = 2S_j^{(z)}\delta_{lj}$$

$$[S_l^{(z)}, S_j^{(+)}] = S_j^{(+)}\delta_{lj}$$

$$[S_l^{(z)}, S_j^{(-)}] = -S_j^{(-)}\delta_{lj} \tag{1.290}$$

The last two commutators can be easily obtained from the direct definition of the operators; for example,

$$[S_j^{(z)}, S_j^{(+)}] = [S_j^{(z)}, S_j^{(x)}] + i[S_j^{(z)}, S_j^{(y)}] = iS_j^{(y)} - i^2 S_j^{(x)} = S_j^{(+)}$$

Spins are neither bosons nor fermions. These commutation relations are unlike any which we have encountered previously. It is precisely these commutation relations which make spin problems so difficult.

1.4.1. Homogeneous Spin Systems

The Heisenberg Hamiltonian puts one spin on each site of a lattice and has the spins interact with a vector interaction. If the interaction is only between nearest neighbor spins, the Hamiltonian has the form

$$H = -J \sum_{j\delta} \mathbf{S}_j \cdot \mathbf{S}_{j+\delta} = -J \sum_{j\sigma} [S_j^{(x)} S_{j+\delta}^{(x)} + S_j^{(y)} S_{j+\delta}^{(y)} + S_j^{(z)} S_{j+\delta}^{(z)}] \tag{1.291}$$

Again the notation $j + \delta$ means the neighbors of site j. The *Heisenberg Hamiltonian* is often solved for spin greater than one-half or for coupling between spins which may be further neighbors. But the above form is the most common. The word solve means "approximately solve," since it cannot be solved exactly, except in one dimension. Often the coupling constant in one direction, say z, is taken to be different from those in the other directions. This case is the anisotropic Heisenberg model:

$$H = -J_\parallel \sum_{j\delta} S_j^{(z)} S_{j+\delta}^{(z)} - J_\perp \sum_{j\delta} [S_j^{(x)} S_{j+\delta}^{(x)} + S_j^{(y)} S_{j+\delta}^{(y)}]$$

$$H = -J_\parallel \sum_{j\delta} S_j^{(z)} S_{j+\delta}^{(z)} - J_\perp \sum_{j\delta} S_j^{(+)} S_{j+\delta}^{(-)} \tag{1.292}$$

In this case the operators $S_j^{(+)}$ and $S_{j+\delta}^{(-)}$ can be arranged in any order. They commute since they refer to different sites. There are two limiting cases of this Hamiltonian which have their own names. The *Ising model* has $J_\perp = 0$

$$H_I = -J_\parallel \sum_{j\delta} S_j^{(z)} S_{j+\delta}^{(z)} \tag{1.293}$$

It may be solved exactly in one dimension, even if one adds a magnetic field to the Hamiltonian:

$$H_I = -J_\parallel \sum_{j\delta} S_j^{(z)} S_{j+\delta}^{(z)} - H_0 \sum_j S_j^{(z)} \tag{1.294}$$

where H_0 is the magnetic field in units of joules. In two dimensions, it may be solved exactly without the magnetic field, as shown by Onsager (1944). Very accurate three-dimensional results have been obtained by a variety of techniques, including Green's functions (Callen, 1966) and critical point and renormalization group techniques (Domb and Green, 1972–1977). The *XY model* has only $J_\perp$:

$$H_{XY} = -J_\perp \sum_{j\delta} S_j^{(+)} S_{j+\delta}^{(-)} \tag{1.295}$$

It may be solved exactly only in one dimension.

It is conventional to write the Hamiltonian in (1.291) with the negative sign in front of the interaction term. Then for $J > 0$, the spins tend to line up all parallel, which is the ferromagnetic arrangement. For $J < 0$, the ordering has alternate spins up and down, if the lattice permits, which is called antiferromagnetic.

The difficulty with solving spin problems is well illustrated by defining collective operators. The operators are transformed into wave vector space:

$$S_{\mathbf{k}}^{(+)} = \frac{1}{\sqrt{N}} \sum_j e^{i\mathbf{k}\cdot\mathbf{R}_j} S_j^{(+)}, \quad S_j^{(+)} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_j} S_{\mathbf{k}}^{(+)} \tag{1.296}$$

$$S_{\mathbf{k}}^{(-)} = \frac{1}{\sqrt{N}} \sum_j e^{-i\mathbf{k}\cdot\mathbf{R}_j} S_j^{(-)}, \quad S_j^{(-)} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_j} S_{\mathbf{k}}^{(-)} \tag{1.297}$$

This transformation appears to be a reasonable approach. It was used successfully in solving fermion and boson problems. But examine the commutation relations for these operators:

$$[S_{\mathbf{k}}^{(+)}, S_{\mathbf{k}'}^{(-)}] = \frac{1}{N} \sum_{jl} e^{i\mathbf{k}\cdot\mathbf{R}_l - i\mathbf{k}'\cdot\mathbf{R}_j} [S_l^{(+)}, S_j^{(-)}]$$

$$[S_{\mathbf{k}}^{(+)}, S_{\mathbf{k}'}^{(+)}] = \frac{2}{N} \sum_j e^{i\mathbf{R}_j\cdot(\mathbf{k}-\mathbf{k}')} S_j^{(z)} \tag{1.298}$$

The operators $S_l^{(+)}$ and $S_j^{(-)}$ commute except on the same site, and then their commutator in (1.290) is $2S_j^{(z)}$. The right-hand side of (1.298) is not simple. It would be preferable to find something like

$$[S_{\mathbf{k}}^{(+)}, S_{\mathbf{k}'}^{(-)}] = C\delta_{\mathbf{k},\mathbf{k}'} \tag{1.299}$$

which would indicate that $S_{\mathbf{k}}^{(+)}$ and $S_{\mathbf{k}'}^{(-)}$ were independent operators except for $\mathbf{k} = \mathbf{k}'$. They would behave just like bosons, and boson statistics would be used. Unfortunately, (1.298) does not have this property. One common approximation (Callen, 1966) is to replace (1.298) by the approximate expression

$$[S_{\mathbf{k}}^{(+)}, S_{\mathbf{k}'}^{(-)}] \approx 2\delta_{\mathbf{k},\mathbf{k}'} \frac{1}{N} \sum_j S_j^{(z)} = 2\delta_{\mathbf{k},\mathbf{k}'} \langle S^{(z)} \rangle \tag{1.300}$$

One then tries to find a self-consistent equation for the average magnetization $\langle S^{(z)} \rangle$, but this approach is approximate. Except in special circumstances, the operators $S_{\mathbf{k}}^{(+)}$ and $S_{\mathbf{k}'}^{(-)}$ do not describe independent eigenstates of the system. This lack of collective eigenstates is the difficulty with solving spin systems and why innocent-looking Hamiltonians like (1.295) are difficult to solve. Hamiltonians with bilinear boson or fermion operators may be solved exactly. That is not true for Hamiltonians with bilinear spin operators.

If the two equations (1.289) are added together, then

$$S_l^{(+)} S_l^{(-)} + S_l^{(-)} S_l^{(+)} = 2[S_l^{(x)2} + S_l^{(y)2}] = 2[\mathbf{S}\cdot\mathbf{S} - S_l^{(z)2}] \tag{1.301}$$

$$= 2[S(S+1) - S_l^{(z)2}] \tag{1.302}$$

For spin one-half, $S_l^{(z)2} = 1/4$, which gives the result

$$\{S_l^{(+)}, S_l^{(-)}\} = 2[\tfrac{1}{2}(\tfrac{3}{2}) - \tfrac{1}{4}] = 1 \tag{1.303}$$

The $S^{(+)}$ and $S^{(-)}$ operators, on the same site, obey an anticommutation relation. Anticommutation relations are associated with fermion operators. The spin one-half operators

behave as a fermion on the same site. For fermions, each site may have zero or one particle for each orbital and spin state. These two possibilities correspond to the two possible spin states of up and down. For example, spin down is equivalent to zero particles and spin up to one particle on the site. Equation (1.288) is analogous to the fermion relations

$$C_j|1\rangle = |0\rangle, \quad C_j^\dagger|0\rangle = |1\rangle$$

$$C_j|0\rangle = 0, \quad C_j^\dagger|1\rangle = 0 \tag{1.304}$$

Unfortunately, one cannot carry this analogy too far. On different sites, the spin operators commute, while the fermion operators anticommute. The spin operators are not fermions either.

However, in one dimension, the spin one-half operators can be made into exact fermion operators. This transformation was discovered by Jordan and Wigner (1928). In one dimension, the spins are aligned along a chain. A new set of operators is defined: the old raising and lowering operators are multiplied by a phase factor which is dependent on spin site:

$$d_j = e^{i\phi_j} S_j^{(-)}$$

$$d_j^\dagger = e^{-i\phi_j} S_j^{(+)}$$

$$d_j^\dagger d_j = S_j^{(+)} S_j^{(-)} = S(S+1) - S_j^{(z)2} + S_j^{(z)} = \tfrac{1}{2} + S_j^{(z)} \tag{1.305}$$

The phase factor ϕ_j is chosen to be π times an operator which measures the number of spin-up operators to the right of that position:

$$j > 1 : \phi_j = \pi \sum_{k=1}^{j-1} (\tfrac{1}{2} + S_k^{(z)}) = \pi \sum_{k=1}^{j-1} d_k^\dagger d_k \tag{1.306}$$

The chain is numbered from one end, say the right, with site indices $j = 1, 2, 3, \ldots$. Set $\phi_1 = 0$. The $d^\dagger$ operators are interpreted as creating fermion particles, and $d_j^\dagger d_j = n_j$ is the number operator for each site. The phase ϕ_j is π times the number of such fermions to the right of the site j. The phase factor commutes with $S_j^{(-)}$, since the operator ϕ_j is the number to the right and does not involve the number operator on the same site. On the same site, these operators anticommute,

$$\{d_j, d_j^\dagger\} = \{S_j^{(-)}, S_j^{(+)}\} = 1 \tag{1.307}$$

since that is the property of the spin operators themselves. On different sites, they also anticommute, with the help of this new phase factor. By taking the anticommutator,

$$\{d_l, d_m^\dagger\} = e^{i\phi_l} S_l^{(-)} e^{-i\phi_m} S_m^{(+)} + e^{-i\phi_m} S_m^{(+)} e^{i\phi_l} S_l^{(-)} \tag{1.308}$$

Since $l \neq m$, assume that $l > m$. The phase factor ϕ_m then commutes with $S_l^{(-)}$, but ϕ_l does not commute with $S_m^{(+)}$. The anticommutator is then

$$\{d_l, d_m^\dagger\} = e^{i\phi_l - i\phi_m} [S_{ell}^{(-)} S_m^{(+)} + e^{-i\pi n_m} S_m^{(+)} e^{i\pi n_m} S_l^{(-)}] \tag{1.309}$$

The right-hand term contains the operator combination

$$e^{-i\pi(1/2+S^{(z)})} S^{(+)} e^{i\pi(1/2+S^{(z)})} = e^{i\pi} S^{(+)} = -S^{(+)} \tag{1.310}$$

The $S^{(+)}$ operator must always raise the magnetic quantum number m by unity, so that $S^{(z)}$ on the left always measures one integer higher value than the same operator on the right of $S^{(+)}$. One gets an extra phase factor of $i\pi$, which changes the sign of the term. This phase factor ϕ_l was chosen to produce this sign change. The anticommutator of the d operators is now

$$\{d_l, d_m^\dagger\} = 0 \quad \text{for } l > m \tag{1.311}$$

It is equal to the commutator of the $S^{(+)}, S^{(-)}$ operators, for different sites, which is zero. The d operators anticommute for both the same site and different sites. They are pure fermion operators and obey fermion statistics. In one dimension the XY model (1.295) may be transformed into a Hamiltonian in terms of the d operators,

$$H_{XY} = -J\sum_j (S_j^{(+)}S_{j+1}^{(-)} + S_j^{(+)}S_{j-1}^{(-)}) \tag{1.312}$$

$$= -J\sum_j (e^{i\phi_j} d_j^\dagger e^{-i\phi_{j+1}} d_{j+1} + e^{i\phi_j} d_j^\dagger e^{-i\phi_{j-1}} d_{j-1}) \tag{1.313}$$

$$H_{XY} = -J\sum_j (d_j^\dagger e^{-i\pi n_j} d_{j+1} + d_j^\dagger e^{-i\pi n_{j-1}} d_{j-1}) \tag{1.314}$$

$$H_{XY} = -J\sum_j (d_j^\dagger d_{j+1} + d_j^\dagger d_{j-1}) \tag{1.315}$$

which is just the tight-binding model for fermions. The phase factors vanish because in the first term n_j is zero if it precedes a raising operator, and in the second term it is zero if it follows a lowering operator. Next the Hamiltonian is changed to collective coordinates:

$$d_j = \frac{1}{\sqrt{N}}\sum_k e^{-ikR_j} d_k, \qquad d_k = \frac{1}{\sqrt{N}}\sum_j e^{ikR_j} d_j$$

$$d_j^\dagger = \frac{1}{\sqrt{N}}\sum_k e^{ikR_j} d_k^\dagger, \qquad d_k^\dagger = \frac{1}{\sqrt{N}}\sum_j e^{-ikR_j} d_j^\dagger$$

$$H = -2J\sum_k \gamma_k d^\dagger d_k, \qquad \gamma_k = \cos(ka) \tag{1.316}$$

and the collective operators now obey anticommutation relations as well:

$$\{d_k, d_{k'}^\dagger\} = \frac{1}{N}\sum_{jl} e^{i(kR_j - k'R_l)}\{d_j, d_l^\dagger\} = \frac{1}{N}\sum_j e^{iR_j(k-k')} = \delta_{kk'} \tag{1.317}$$

The Hamiltonian may be written as a simple fermion problem. The exact partition function is

$$Z = \mathrm{Tr}e^{-\beta H} = \Pi_k(1 + e^{2J\gamma_k\beta}) \tag{1.318}$$

Note that there is no chemical potential for spin systems. One can also work out other properties of this model. To prove that the transformation (1.305) is valid, one should show that all the commutation relations (1.290) are preserved. They are.

The Jordan–Wigner transformation shows that in one dimension the spin one-half operators may be represented exactly as fermions. This result is not valid for higher dimensions. No one has been able to find an equivalent transformation for two or three

dimensions. Indeed, most approximate analyses assume that in two or three dimensions the spin excitations behave approximately as bosons rather than as fermions.

In two and three dimensions, one may still transform spin one-half operators into particle operators. In this case, the particle operators have funny commutation relations. They are, like spins, neither fermions nor bosons. This transformation associates the creation operator with $S^{(+)}$ and a destruction operator with $S^{(-)}$:

$$C_l^\dagger = S_l^{(+)}$$
$$C_l = S_l^{(-)} \tag{1.319}$$
$$n_l = C_l^\dagger C_l = S_l^{(+)} S_l^{(-)} = \tfrac{1}{2} + S_l^{(z)} \tag{1.320}$$

This transformation preserves all the commutation relations (1.290). For example, one has that

$$\{C_l, C_l^\dagger\} = 1 \tag{1.321}$$
$$[C_l, C_m^\dagger] = -2(\tfrac{1}{2} - n_l)\delta_{lm} \tag{1.322}$$
$$[C_l, n_l] = [S_l^{(-)}, S_l^{(z)}] = C_l \tag{1.323}$$

The particle operators C_l and $C_m^\dagger$ anticommute if they are on the same site and commute if they are on different sites. They are neither fermions nor bosons. Collective operators such as

$$C_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_j e^{i\mathbf{k}\cdot\mathbf{r}_j} C_j \tag{1.324}$$

have funny commutation relations, similar to the spin case in (1.298). Nevertheless, this Hamiltonian is a popular many-body model for certain systems. These are *lattice gas models* for atoms on lattices. The atoms may be considered as classical particles, which commute on different sites. However, there may not be more than one atom on each site, since the atoms are large, substantial objects. "No more than one atom on each site" is an exclusion principle, which is represented by the anticommutation relations on the same site. The same physics is contained in a model which has the particle obey purely boson statistics but with the provision that there is a strong repulsive interaction U if two particles were on the same atomic site. The lattice gas results would be obtained in the limit $U \to \infty$. The statistics of anticommutation relations on the same site merely represent the strong repulsive interaction between atoms at close separation.

The usual *lattice gas* (LG) model has pairwise interactions U between particles in nearest neighbor positions. A chemical potential μ is also introduced for the particles, since they may have variable concentration:

$$K_{LG} = H_{LG} - \mu N = \tfrac{1}{2} U \sum_{j\delta} n_j n_{j+\delta} - \mu \sum_j n_j \tag{1.325}$$

The factor of one-half is due to counting each pair twice in the summation over all $(j\delta)$. The lattice gas Hamiltonian may be transformed into an equivalent magnetic problem by inverting the transformation of (1.319):

$$K_{LG} = \frac{1}{2}U\sum_{j\delta}\left(S_j^{(z)} + \frac{1}{2}\right)\left(S_{j+\delta}^{(z)} + \frac{1}{2}\right) - \mu\sum_j\left(S_j^{(z)} + \frac{1}{2}\right) \tag{1.326}$$

$$K_{LG} = \frac{1}{2}U\sum_{j\delta}S_j^{(z)}S_{j+\delta}^{(z)} + \left(\frac{U\mathcal{Z}}{2} - \mu\right)\sum_j S_j^{(z)} - \frac{1}{2}N_0\left(\frac{U\mathcal{Z}}{4} - \mu\right) \tag{1.327}$$

where $\mathcal{Z}$ is the coordination number, i.e., the number of nearest neighbors. The spin version of K_{LG} is identical to the Ising model with magnetic field. The magnetic field is $H_0 = U\mathcal{Z}/2 - \mu$. If exactly one-half of the sites of the lattice gas are occupied, then one has $\mu = U\mathcal{Z}/2$ so that the effective magnetic field is zero (Hill, 1956). In this case, for a half-filled band, the chemical potential is temperature independent. For concentrations other than one-half, the chemical potential varies with temperature.

The *quantum lattice gas* (QLG) model adds a nearest neighbor hopping term to the lattice gas model:

$$K_{QLG} = H_{QLG} - \mu N = \frac{1}{2}U\sum_{j\delta}n_j n_{j+\delta} + w\sum_{j\delta}C_j^\dagger C_{j+\delta} - \mu\sum_j n_j \tag{1.328}$$

One can show that the equivalent magnetic problem is the anisotropic Heisenberg model with magnetic field. The quantum lattice gas was suggested by Matsubara and Matsuda (1956) as a model for quantum fluids such as ^{4}He. The superfluid transformation for this system occurs in the liquid state. Nevertheless, a quantum lattice gas model appears to be a good description of its critical properties. The parameter U may be taken to be either positive or negative, depending on whether the nearest neighbors repulse or attract each other. The equivalent magnetic problems are then antiferromagnetic and ferromagnetic, respectively.

1.4.2. Impurity Spin Models

The models mentioned so far are for homogeneous magnetic systems. The same spin was on each site, and we tried to deduce the magnetic properties of the entire system. Other kinds of popular models are for impurity spin problems. Here the spin is an isolated impurity in an otherwise homogeneous electron gas. One can study, for example, the conditions for the formation of a local moment on the impurity or the scattering properties of the free electrons from the localized spin.

The derivation follows Kondo (1969). An impurity atom is located at the position $\mathbf{R}_n$ and has a localized electron orbital $\phi_L(\mathbf{r} - \mathbf{R}_n)$ when the electron is on that site. Otherwise the electron is in a continuum state $\mathbf{k}$ with wave function $\phi_k(\mathbf{r})$ and energy $\varepsilon_\mathbf{k}$. The wave functions may be considered as plane waves or alternately as Bloch functions of the crystal. A generalized state function is the summation over all possible states,

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}\sigma}\phi_\mathbf{k}(\mathbf{r})X_\sigma C_{\mathbf{k}\sigma} \tag{1.329}$$

where the X_σ are the spin wave functions, which denote spin up $X_\uparrow$ or down $X_\downarrow$. The wave functions $\phi_\mathbf{k}(\mathbf{r})$ and $\phi_L(\mathbf{r} - \mathbf{R}_\mathbf{n})$ are not assumed to be orthogonal. This step will be taken

later, since their orthogonality is a many-body problem of sorts. The Hamiltonian is taken to have terms such as

$$H = \sum_i \left[\frac{p_i^2}{2m} + U(\mathbf{r}_i)\right] + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{r_{ij}} \tag{1.330}$$

where $U(\mathbf{r})$ is the potential for each electron, which may include the impurity potential as well as the usual potential of the host lattice. The last term is the electron–electron interactions. The first terms of the Hamiltonian are evaluated for the state function, and one gets

$$\begin{aligned}\int d^3r\psi^\dagger(\mathbf{r})\left[\frac{p_i^2}{2m} + U(\mathbf{r}_i)\right]\psi(\mathbf{r}) = &\sum_{\mathbf{k}\sigma}\varepsilon_{\mathbf{k}}C^\dagger_{\mathbf{k}\sigma}C_{\mathbf{k}\sigma} + \sum_\sigma \varepsilon_L C^\dagger_{L\sigma}C_{L\sigma}\\ &+ \sum_{\mathbf{k}\mathbf{k}'\sigma} U_{\mathbf{k}\mathbf{k}'}C^\dagger_{\mathbf{k}'\sigma}C_{\mathbf{k}\sigma}\\ &+ \sum_{\mathbf{k}\sigma} M_{\mathbf{k}}(C^\dagger_{\mathbf{k}\sigma}C_{L\sigma} + C^\dagger_{L\sigma}C_{\mathbf{k}\sigma})\end{aligned} \tag{1.331}$$

The first two terms on the right are the unperturbed parts of the Hamiltonian for the continuum states and the impurity. The impurity states are labeled $(L\sigma)$. The third term is

$$\sum_{\mathbf{k}\mathbf{k}'\sigma} U_{\mathbf{k}\mathbf{k}'}C^\dagger_{\mathbf{k}'\sigma}C_{\mathbf{k}\sigma} \tag{1.332}$$

which involves the scattering of the continuum functions from the impurity potential. This problem is simple to solve by ordinary scattering theory. The solution will be given in Sec. 4.1. The last term is a mixing term called H_M:

$$H_M = \sum_{\mathbf{k}\sigma} M_{\mathbf{k}}(C^\dagger_{\mathbf{k}\sigma}C_{L\sigma} + C^\dagger_{L\sigma}C_{\mathbf{k}\sigma}) \tag{1.333}$$

The interaction H_M is sometimes called H_{sd} but that name is used for another contribution which will be discussed shortly. The mixing term describes processes whereby the electron hops off of the impurity and becomes a continuum state or vice versa. This term essentially arises from the nonorthogonality of the continuum and local wave functions. Note that the spin direction is maintained during this hybridization process. A Hamiltonian of the type

$$H = \sum_{\mathbf{k}\sigma}\varepsilon_{\mathbf{k}}C^\dagger_{\mathbf{k}\sigma}C_{\mathbf{k}\sigma} + \sum_\sigma \varepsilon_L C^\dagger_{L\sigma}C_{L\sigma} + \sum_{\mathbf{k}\sigma} M_{\mathbf{k}}(C^\dagger_{\mathbf{k}\sigma}C_{L\sigma} + C^\dagger_{L\sigma}C_{\mathbf{k}\sigma}) \tag{1.334}$$

will be called a *Fano–Anderson model*, since it was introduced simultaneously by Fano (1961) and Anderson (1961). It should not be confused with the famous Anderson model, which will be described below. The Fano–Anderson model may be solved exactly, and this solution will be presented in Sec. 4.2. There is no real conceptual difference between (1.331) and (1.334). If the wave functions $\psi_{\mathbf{k}}(\mathbf{r})$ are chosen to be eigenstates of the Hamiltonian which includes the impurity scattering $U'_{\mathbf{k}\mathbf{k}}$, then the two Hamiltonians become identical.

The interesting magnetic phenomenon comes from the terms involving electron–electron interactions. Consider

$$\int d^3r_1 d^3r_2 \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}\psi^\dagger(\mathbf{r}_1)\psi(\mathbf{r}_1)\psi^\dagger(\mathbf{r}_2)\psi(\mathbf{r}_2) \tag{1.335}$$

Many types of terms are generated by this expression. Only two of them are discussed here. The first contains operators from two continuum functions and two localized wave functions:

$$C^\dagger_{L\sigma} C_{L\sigma'} C^\dagger_{\mathbf{k}\sigma''} C_{\mathbf{k}'\sigma'''} \tag{1.336}$$

The four spin operators must occur in pairs. One possible pairing ($\sigma = \sigma'$, $\sigma'' = \sigma'''$) gives the interaction

$$\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} V_{\mathbf{k}\mathbf{k}'} C^\dagger_{L\sigma} C_{L\sigma} C^\dagger_{\mathbf{k}\sigma'} C_{\mathbf{k}'\sigma'} \tag{1.337}$$

$$V_{\mathbf{k}\mathbf{k}'} = e^2 \int d^3r_1 \phi^*_{\mathbf{k}}(\mathbf{r}_1)\phi_{\mathbf{k}'}(\mathbf{r}_1) \int \frac{d^3r_2}{|\mathbf{r}_1 - \mathbf{r}_2|} |\phi_L(\mathbf{r}_2 - \mathbf{R}_n)|^2 \tag{1.338}$$

This term is usually ignored, since it does not cause magnetic phenomena. It states that the conduction electrons interact with the impurity in a different way when the localized orbital is occupied. It is quite a reasonable term, since the electron is charged and one expects that the presence of this charge will influence the other electrons. This term is customarily ignored. Another possible spin arrangement is the exchange term for the above process, which has the form

$$-\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} J_{\mathbf{k}\mathbf{k}'} C^\dagger_{L\sigma'} C_{L\sigma} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{k}'\sigma'}$$

$$J_{\mathbf{k}\mathbf{k}'} = e^2 \int d^3r_1 \phi^*_{\mathbf{k}}(\mathbf{r}_1)\phi_L(\mathbf{r}_1 - \mathbf{R}_n) \int \frac{d^3r_2}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi^*_L(\mathbf{r}_2 - \mathbf{R}_n)\phi_{\mathbf{k}'}(\mathbf{r}_2) \tag{1.339}$$

In this term, the electron which is scattering may change its spin state during the scattering process. This spin change is always accompanied by an opposite spin change of the impurity spins, so that the total spin angular momentum is conserved during the process. The spin conservation is illustrated by writing out in detail the terms which can occur:

$$C^\dagger_{\mathbf{k}\uparrow} C_{\mathbf{k}'\uparrow} C^\dagger_{L\uparrow} C_{L\uparrow} + C^\dagger_{\mathbf{k}\uparrow} C_{\mathbf{k}'\downarrow} C^\dagger_{L\downarrow} C_{L\uparrow} + C^\dagger_{\mathbf{k}\downarrow} C_{\mathbf{k}'\uparrow} C^\dagger_{L\uparrow} C_{L\downarrow} + C^\dagger_{\mathbf{k}\downarrow} C_{\mathbf{k}'\downarrow} C^\dagger_{L\downarrow} C_{L\downarrow}$$

The first and last terms are regrouped as

$$\begin{aligned} &\tfrac{1}{2}(C^\dagger_{\mathbf{k}\uparrow} C_{\mathbf{k}'\uparrow} + C^\dagger_{\mathbf{k}\downarrow} C_{\mathbf{k}'\downarrow})(C^\dagger_{L\uparrow} C_{L\uparrow} + C^\dagger_{L\downarrow} C_{L\downarrow}) \\ &\quad + \tfrac{1}{2}(C^\dagger_{\mathbf{k}\uparrow} C_{\mathbf{k}'\uparrow} - C^\dagger_{\mathbf{k}\downarrow} C_{\mathbf{k}'\downarrow})(C^\dagger_{L\uparrow} C_{L\uparrow} - C^\dagger_{L\downarrow} C_{L\downarrow}) \end{aligned} \tag{1.340}$$

The first term in (1.340) contains the factor

$$C^\dagger_{L\uparrow} C_{L\uparrow} + C^\dagger_{L\downarrow} C_{L\downarrow} = \sum_\sigma C^\dagger_{L\sigma} C_{L\sigma} \tag{1.341}$$

which is always unity if the localized state is occupied. In fact this first term has the form

$$\frac{1}{2} \sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} J_{\mathbf{k}\mathbf{k}'} C^\dagger_{L\sigma} C_{L\sigma} C^\dagger_{\mathbf{k}\sigma'} C_{\mathbf{k}'\sigma'} \tag{1.342}$$

which is exactly the same as (1.337). The usual convention is to combine these terms, and then ignore both of them. The second term in (1.340) has the combination

$$S^{(z)}_L = \tfrac{1}{2}(C^\dagger_{L\uparrow} C_{L\uparrow} - C^\dagger_{L\downarrow} C_{L\downarrow}) \tag{1.343}$$

which is just the z component of the localized spin. Similarly, some of the other combinations are identified as the raising and lowering operators for the localized spin:

$$S_L^{(+)} = C_{L\uparrow}^\dagger C_{L\downarrow} \tag{1.344}$$

$$S_L^{(-)} = C_{L\downarrow}^\dagger C_{L\uparrow} \tag{1.345}$$

The spin-dependent terms in (1.339) may be collected as

$$H_{sd} = -\sum_{\mathbf{k}\mathbf{k}'} J_{\mathbf{k}\mathbf{k}'}[S_L^{(z)}(C_{\mathbf{k}\uparrow}^\dagger C_{\mathbf{k}\uparrow} - C_{\mathbf{k}\downarrow}^\dagger C_{\mathbf{k}\downarrow}) + S_L^{(+)} C_{\mathbf{k}\downarrow}^\dagger C_{\mathbf{k}'\uparrow} + S_L^{(-)} C_{\mathbf{k}\uparrow}^\dagger C_{\mathbf{k}'\downarrow}]$$

This interaction is called H_{sd}. It is a model of the localized d-electrons interacting with the s-like continuum wave functions. This form of the interaction is valid even if the localized spin has $S > \frac{1}{2}$. The last two terms flip the spin of the continuum electron while flipping the localized spin of the impurity in the opposite direction. The first term does not flip spin, but the interaction does depend on the z-component of the spin of both the impurity and continuum wave function. A Hamiltonian of the type

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} C_{\mathbf{k}\sigma}^\dagger C_{\mathbf{k}\sigma} + \varepsilon_L(n_{L\uparrow} + n_{L\downarrow}) + H_{sd} \tag{1.346}$$

is called the *Kondo problem*. It was not formulated by Kondo (1964) but rather much earlier by Zener (1951). Kondo's contribution was recognizing that the spin-flip scattering processes could cause unusual low-temperature behavior in the scattering properties. These low-temperature anomalies had long been observed in resistivities (Gerritsen and Linde, 1951, 1954). They are derived in Chapter 6. Another type of term which may arise from the electron–electron interaction (1.335) has four local operators:

$$UC_{L\sigma}^\dagger C_{L\sigma'}^\dagger C_{L\sigma''} C_{L\sigma'''} \tag{1.347}$$

$$U = \int d^3r_1 d^3r_2 |\phi_L(\mathbf{r}_1 - \mathbf{R}_n)|^2 \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} |\phi_L(\mathbf{r}_2 - \mathbf{R}_n)|^2 \tag{1.348}$$

If the orbital is nondegenerate, as in an s-orbital, then two electrons can be on the same site only in opposite spin states. The spin indices must pair up as ($\sigma''' = \sigma$, $\sigma'' = \sigma'$). One gets a term similar to the one in the Hubbard model,

$$Un_{L\uparrow} n_{L\downarrow} \tag{1.349}$$

although it was historically introduced first in the *Anderson model*. This famous model considers the model Hamiltonian

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} C_{\mathbf{k}\sigma}^\dagger C_{\mathbf{k}\sigma} + \varepsilon_L \sum_\sigma C_{L\sigma}^\dagger C_{L\sigma}$$
$$+ \sum_{\mathbf{k}\sigma} M_{\mathbf{k}}(C_{\mathbf{k}\sigma}^\dagger C_{L\sigma} + C_{L\sigma}^\dagger C_{\mathbf{k}\sigma}) + Un_{L\uparrow} n_{L\downarrow} \tag{1.350}$$

The Hamiltonian is exactly solvable without the last term. The last term is very important, since it causes magnetic instabilities in some circumstances. The Hamiltonian describes the interaction of the conduction electrons with a single impurity at site $\mathbf{R}_n$. The index L denotes the orbital state of the impurity. If there is only one important orbital state, as is the case for s-states, then the indices are often dropped from the notation.

The Anderson model and the Kondo model both describe the interaction of a continuum electron with a localized one. The two models are not totally different. There is a canonical transformation which, when applied to the Anderson model, will transform it into a form similar to the Kondo model. This transformation on the Anderson model produces quite a few terms, of which the Kondo model is a subset. The transformation does not produce exactly the Kondo model, and the two models are not identical. Write the Anderson model as $(C_{Ls} \equiv C_s, n_s \equiv C_s^\dagger C_s)$

$$H = H_0 + H_M \tag{1.351}$$

$$H_0 = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} C_{\mathbf{k}\sigma}^\dagger C_{\mathbf{k}\sigma} + \varepsilon_L \sum_\sigma C_{L\sigma}^\dagger C_{L\sigma} + U n_{L\uparrow} n_{L\downarrow} \tag{1.352}$$

$$H_M = \sum_{\mathbf{k}\sigma} M_{\mathbf{k}} (C_{\mathbf{k}\sigma}^\dagger C_{L\sigma} + C_{L\sigma}^\dagger C_{\mathbf{k}\sigma}) \tag{1.353}$$

then a canonical transformation (Schrieffer and Wolff, 1966) is

$$\begin{aligned}\hat{H} = e^S H e^{-S} &= H_0 + H_M + [S, H_0] + [S, H_M] + \tfrac{1}{2}[S, [S, H_0]] \\ &+ \tfrac{1}{2}[S, [S, H_M]] + \cdots\end{aligned} \tag{1.354}$$

The function S is chosen so that all terms are eliminated which are linear in $M_{\mathbf{k}}$. It is accomplished by choosing S to be

$$0 = H_M + [S, H_0] \tag{1.355}$$

and then the canonical transformation produces the series

$$\hat{H} = H_0 + \tfrac{1}{2}[S, H_M] + \tfrac{1}{3}[S, [S, H_M]] + \cdots \tag{1.356}$$

There are an infinite number of terms in the series. Only those which are proportional to $M_{\mathbf{k}}^2$ will be evaluated here, which is the second term in the series (1.356). The transformation factor S turns out to be

$$S = \sum_{\mathbf{k}s} M_{\mathbf{k}} \left(\frac{1 - n_{-s}}{\varepsilon_L - \varepsilon_{\mathbf{k}}} + \frac{n_{-s}}{\varepsilon_L + U - \varepsilon_{\mathbf{k}}} \right) (C_s^\dagger C_{\mathbf{k}s} - C_{\mathbf{k}s}^\dagger C_s) \tag{1.357}$$

so that one gets the commutators

$$\left[S, \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} n_{\mathbf{k}\sigma} \right] = \sum_{\mathbf{k}s} M_{\mathbf{k}} \left(\frac{1 - n_{-s}}{\varepsilon_L - \varepsilon_{\mathbf{k}}} + \frac{n_{-s}}{\varepsilon_L + U - \varepsilon_{\mathbf{k}}} \right) (C_s^\dagger C_{\mathbf{k}s} + C_{\mathbf{k}s}^\dagger C_s) \varepsilon_{\mathbf{k}}$$

$$\left[S, \varepsilon_L \sum_s n_s \right] = -\varepsilon_L \sum_{\mathbf{k}s} M_{\mathbf{k}} \left(\frac{1 - n_{-s}}{\varepsilon_L - \varepsilon_{\mathbf{k}}} + \frac{n_{-s}}{\varepsilon_L + U - \varepsilon_{\mathbf{k}}} \right) (C_s^\dagger C_{\mathbf{k}s} + C_{\mathbf{k}s}^\dagger C_s)$$

$$[S, U n_\uparrow n_\downarrow] = -U \sum_{\mathbf{k}s} M_{\mathbf{k}} \left(\frac{1 - n_{-s}}{\varepsilon_L - \varepsilon_{\mathbf{k}}} + \frac{n_{-s}}{\varepsilon_L + U - \varepsilon_{\mathbf{k}}} \right) (C_s^\dagger C_{\mathbf{k}s} + C_{\mathbf{k}s}^\dagger C_s) n_{-s}$$

When these are added, one exactly satisfies (1.355). Although the form of S appears unwieldly, it gets the job done. The next step is to take the commutator $[S, H_M]$ to generate the

terms which are proportional to $M_{\mathbf{k}}^2$. As a preliminary step, define the following effective exchange constant:

$$J_{\mathbf{k}\mathbf{k}'} = M_{\mathbf{k}} M_{\mathbf{k}'} \left(\frac{1}{\varepsilon_L - \varepsilon_{\mathbf{k}}} + \frac{1}{\varepsilon_L - \varepsilon_{\mathbf{k}'}} - \frac{1}{\varepsilon_L + U - \varepsilon_{\mathbf{k}}} - \frac{1}{\varepsilon_L + U - \varepsilon'_{\mathbf{k}}} \right)$$

In terms of this constant, one finds the following for the commutator:

$$\begin{aligned}
[S, H_M] = \sum_{\mathbf{k}\mathbf{k}'s} J_{\mathbf{k}\mathbf{k}'} [& n_{-s} C_{\mathbf{k}s}^\dagger C_{\mathbf{k}'s} + C_{\mathbf{k}'s}^\dagger C_{\mathbf{k}-s} C_{-s}^\dagger C_s - \delta_{\mathbf{k}\mathbf{k}'} n_s n_{-s} \\
& + \tfrac{1}{4} (C_{\mathbf{k}-s}^\dagger C_{\mathbf{k}'s}^\dagger C_{-s} C_s + \text{h.c.})] + 2 \sum_{\mathbf{k}} \frac{M_{\mathbf{k}}^2}{\varepsilon_L - \varepsilon_{\mathbf{k}}} n_s \\
& - \sum_{\mathbf{k}\mathbf{k}'s} C_{\mathbf{k}s}^\dagger C_{\mathbf{k}'s} M_{\mathbf{k}} M_{\mathbf{k}'} \left(\frac{1}{\varepsilon_L - \varepsilon_{\mathbf{k}}} + \frac{1}{\varepsilon_L - \varepsilon_{\mathbf{k}'}} \right)
\end{aligned} \tag{1.358}$$

The second term flips the spin of the electron and impurity electron while scattering—it has exactly the Kondo form. There are many other terms. The last term is potential scattering of the continuum electron, which must be added to our earlier result (1.331). There is a term which is the non-spin-flip interaction between a continuum electron and an impurity electron and this must be added to (1.337). There is also a term which contains the interaction between two impurity electrons, one with spin up and the other with spin down. This renormalizes U in (1.349). Finally, there is a new term which has not been previously encountered. It describes the process

$$C_{\mathbf{k},-s}^\dagger C_{\mathbf{k}'s}^\dagger C_{-s} C_s \tag{1.359}$$

whereby two impurity electrons hop off of the impurity site to become two continuum electrons and vice versa. The Anderson model certainly describes a rich set of phenomena. Of course, many more terms are generated by the additional commutators in the series (1.356).

1.5. PHOTONS

Throughout this book the following type of Hamiltonian will be used for the discussion of the interaction of charges e_j with each other and with a radiation field. The raising ($a_{\mathbf{k}\lambda}^\dagger$) and lowering ($a_{\mathbf{k}\lambda}$) operators are for the photons:

$$H = \sum_i \frac{1}{2m} \left[\mathbf{p}_i - \frac{e_i}{c} \mathbf{A}(\mathbf{r}_i) \right]^2 + \frac{1}{2} \sum_{i \neq j} \frac{e_i e_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\mathbf{k}\lambda} \omega_{\mathbf{k}} a_{\mathbf{k}\lambda}^\dagger a_{\mathbf{k}\lambda} \tag{1.360}$$

The radiation field is represented by the vector potential

$$\frac{1}{c} A_\mu(\mathbf{r}, t) = \sum_{\mathbf{k}\lambda} \left(\frac{2\pi}{\nu \omega_{\mathbf{k}}} \right)^{1/2} \xi_\mu(\mathbf{k}, \lambda) (a_{\mathbf{k}\lambda} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega_{\mathbf{k}} t)} + \text{h.c.}) \tag{1.361}$$

$$= \frac{1}{\sqrt{\nu}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} A_\mu(\mathbf{k}, t) \tag{1.362}$$

The term $\sum \omega_k a^\dagger_{k\lambda} a_{k\lambda}$ represents its unperturbed photon Hamiltonian in the absence of charges. The unit polarization vector is ξ_μ. One feature of this Hamiltonian is the term

$$\frac{1}{2}\sum_{i\neq j}\frac{e_i e_j}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.363}$$

which is the Coulomb interaction between charges. The Coulomb interaction is instantaneous in time, since the potential has no retardation, or speed of light, built into it. The lack of retardation is not an approximation but is rigorously correct in the Coulomb gauge. Of course, there is retardation in the total interaction, which arises through the vector potential fields.

1.5.1. Gauges

Although the topic of gauges is treated correctly in a number of texts, it still seems to be poorly understood by students. It seems appropriate to start at the beginning and reproduce some standard material. Maxwell equations are

$$\nabla \cdot \mathbf{B} = 0 \tag{1.364}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{1.365}$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho \tag{1.366}$$

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j} \tag{1.367}$$

The charge density is $\rho(\mathbf{r}, t)$ and the current density is $\mathbf{j}(\mathbf{r}, t)$. There is an important theorem that any vector function of position can be written as the sum of two terms: one is the gradient of a potential, and the other is the curl of a vector:

$$\mathbf{S}(\mathbf{r}) = \nabla g + \nabla \times \mathbf{m}(\mathbf{r}) = \mathbf{S}_l + \mathbf{S}_t \tag{1.368}$$

$$\mathbf{S}_l = \nabla g \tag{1.369}$$

$$\mathbf{S}_t = \nabla \times \mathbf{m} \tag{1.370}$$

The term $\mathbf{S}_l$ is called the *longitudinal* part of $\mathbf{S}$, and $\mathbf{S}_t$ is called the *transverse* part. If $\mathbf{B}(\mathbf{r})$ is assumed to have this form, then Eq. (1.364) becomes

$$\nabla \cdot (\nabla g + \nabla \times \mathbf{A}) = \nabla^2 g = 0 \tag{1.371}$$

Usually $g = 0$, so that the vector potential is defined as

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{1.372}$$

However, this expression does not uniquely define $\mathbf{A}(\mathbf{r})$. The definition $\mathbf{B} = \nabla \times \mathbf{A}$ is put into (1.365):

$$\nabla \times \left[\mathbf{E} + \frac{1}{c}\frac{\partial}{\partial t}\mathbf{A}\right] = 0 \tag{1.373}$$

Now the factor in brackets is also the sum of a longitudinal and transverse part:

$$\nabla \times [\nabla\psi + \nabla \times \mathbf{M}] = 0 = \nabla \times (\nabla \times \mathbf{M}) \tag{1.374}$$

The equation is satisfied if $\mathbf{M} = 0$, so that for the electric field is

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \nabla\psi \tag{1.375}$$

where ψ is the scalar potential. When these two forms for $\mathbf{B}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$ are put into (1.366) and (1.367), the equations for the scalar and vector potentials are:

$$\nabla^2\psi + \frac{1}{c}\frac{\partial}{\partial t}\nabla\cdot\mathbf{A} = -4\pi\rho \tag{1.376}$$

$$\nabla\times(\nabla\times\mathbf{A}) + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} + \nabla\frac{\partial}{c\partial t}\psi = \frac{4\pi}{c}\mathbf{j} \tag{1.377}$$

At first it appears that there are four equations for the four unknowns (A_x, A_y, A_z, ψ): Eq. (1.376) and the three vector components of (1.377). If this assertion were true, the four unknown functions would be determined uniquely in terms of the sources $(\mathbf{j}, \rho)$. However, these four equations are not linearly independent; only three of them are independent. The linear dependence is shown by operating on (1.376) by $(1/c)(\partial/\partial t)$ and on (1.377) by ∇. Then subtract the two equations and find

$$\nabla\cdot[\nabla\times(\nabla\times\mathbf{A})] = -\frac{4\pi}{c}\left(\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j}\right) \tag{1.378}$$

The left-hand side is zero because it is the gradient of a curl. The right-hand side vanishes since it is the equation of continuity. The four equations are not independent. There are only three equations for the four unknowns.

Therefore, the four unknown functions (A_x, A_y, A_z, ψ) are not uniquely determined. It is necessary to stipulate one additional condition, or constraint, on their values. It is called the *gauge condition*. The imposed condition is that the Coulomb field $\psi(\mathbf{r}, t)$ shall act instantaneously, which is accomplished by insisting that

$$\nabla\cdot\mathbf{A} = 0 \tag{1.379}$$

Equation (1.379) defines the *Coulomb gauge*, sometimes called the *transverse gauge*. The latter name arises because (1.379) implies that $\nabla\cdot\mathbf{A}_l = 0$, so that $\mathbf{A}$ is purely transverse. One should realize that any arbitrary constraint may be imposed as long as one can satisfy (1.376) and (1.377). As long as these two equations are satisfied, one always obtains the same value for $\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r})$. The arbitrary choice of gauge does not alter the final value of observable quantities.

In the Coulomb gauge Eq. (1.376) simplifies to

$$\nabla^2\psi = -4\pi\rho \tag{1.380}$$

which is easily solved to give

$$\psi(\mathbf{r}, t) = \int d^3r' \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} \tag{1.381}$$

The potential $\psi(\mathbf{r}, t)$ is instantaneous and is not retarded. This result is not an approximation but is an exact consequence for our choice of gauge. Later it is shown that a different choice of gauge leads to a retarded scalar potential.

Next the other equation (1.377) is evaluated. The following identity is useful:

$$\nabla\times(\nabla\times\mathbf{A}) = -\nabla^2\mathbf{A} + \nabla(\nabla\cdot\mathbf{A}) \tag{1.382}$$

The second term vanishes in the Coulomb gauge, which gives

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = -\frac{4\pi}{c}\mathbf{j} + \nabla\frac{\partial\psi}{c\partial t} \tag{1.383}$$

It is useful to operate a bit on the second term on the right-hand side. Using (1.381), this term is

$$\nabla\frac{\partial\psi}{c\partial t} = \frac{1}{c}\nabla\int d^3r'\,\frac{(\partial/\partial t)\rho(\mathbf{r}', t)}{|\mathbf{r}-\mathbf{r}'|} \tag{1.384}$$

$$\nabla\frac{\partial\psi}{c\partial t} = -\frac{1}{c}\nabla\int d^3r'\,\frac{1}{|\mathbf{r}-\mathbf{r}'|}\nabla'\cdot\mathbf{j}(\mathbf{r}', t) \tag{1.385}$$

where the last identity uses the equation of continuity. Integrate by parts in the last term,

$$\nabla\frac{\partial\psi}{c\partial t} = \frac{1}{c}\nabla\int d^3r'\mathbf{j}(\mathbf{r}', t)\cdot\nabla'\frac{1}{|\mathbf{r}-\mathbf{r}'|} \tag{1.386}$$

and then pull the gradient out by letting it operate on $\mathbf{r}$ instead of $\mathbf{r}'$; the latter step requires a sign change. Also operate on the current term itself in (1.383) by using the identity

$$4\pi\mathbf{j}(\mathbf{r}, t) = -\nabla^2\int d^3r'\,\frac{\mathbf{j}(\mathbf{r}', t)}{|\mathbf{r}-\mathbf{r}'|} \tag{1.387}$$

By combining these results

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = (\nabla^2 - \nabla\nabla)\frac{1}{c}\int d^3r'\,\frac{\mathbf{j}(\mathbf{r}', t)}{|\mathbf{r}-\mathbf{r}'|} \tag{1.388}$$

Finally, using the identity (1.382),

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = -\frac{1}{c}\nabla\times\left[\nabla\times\int d^3r'\,\frac{\mathbf{j}(\mathbf{r}', t)}{|\mathbf{r}-\mathbf{r}'|}\right] \tag{1.389}$$

The point of this exercise is that the right-hand side of (1.389) is now a transverse vector, since it is the curl of something. The current is written as a longitudinal plus a transverse part,

$$\mathbf{j} = \mathbf{j}_l + \mathbf{j}_t \tag{1.390}$$

Then the vector potential obeys the equation

$$\nabla^2\mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = -\frac{4\pi}{c}\mathbf{j}_t(\mathbf{r}, t) \tag{1.391}$$

where $\mathbf{j}_t(\mathbf{r}, t)$ is defined by the right-hand side of (1.389). The final equation for $\mathbf{A}$ is very reasonable. Since the vector potential $\mathbf{A}$ is purely transverse, it should respond only to the transverse part of the current. If it were to respond to the longitudinal part of the current, it would develop a longitudinal part. The longitudinal component of $\mathbf{A}$ does not occur in the Coulomb gauge.

As a simple example, consider a current of the form

$$\mathbf{j}(\mathbf{r}) = \mathbf{J}_0 e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.392}$$

whose transverse part is

$$\mathbf{j}_t = -\frac{1}{k^2}\mathbf{k}\times(\mathbf{k}\times\mathbf{J}_0)e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.393}$$

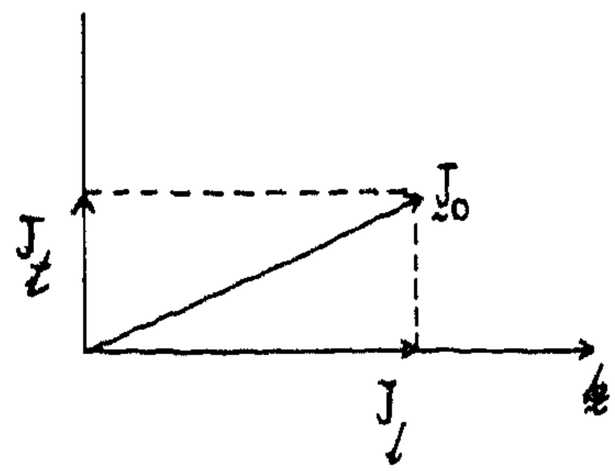


FIGURE 1.2

and $\mathbf{j}_t$ is the component of $\mathbf{j}$ which is perpendicular to $\mathbf{k}$; see Fig. 1.2. In free space, the transverse and longitudinal parts are just those components which are perpendicular and parallel to the wave vector $\mathbf{k}$ of the photon. However, solids are not homogeneous but periodic. Along major symmetry directions in the crystal, it is often true that transverse components are perpendicular to the wave vector. However, it is generally not true for arbitrary points in the Brillouin zone, even in cubic crystals. The words *transverse* and *longitudinal* do not necessarily mean perpendicular and parallel to $\mathbf{k}$.

Two charges interact by the sum of the two interactions: Coulomb plus photon. The net interaction may not have a component which is instantaneous. In fact, for a frequency-dependent charge density, at distances large compared to c/ω, one finds that the photon part of the interaction produces a term $-e^2/r$ which exactly cancels the instantaneous Coulomb interaction. The remaining parts of the photon contribution are the net retarded interaction. Solid state physics is usually concerned with interaction over short distances. Then retardation is unimportant for most problems. In the study of the homogeneous electron gas, for example, the photon part is small and may be neglected. In real solids, the photon part causes some crystal field effects, which is an unexciting many-body effect. The main effect of retardation is the polariton effects at long wavelength (Hopfield, 1958). In general, the Coulomb gauge is chosen because the instantaneous Coulomb interaction is usually a large term which forms a central part of the analysis, while the photon parts are usually secondary. Like most generalizations, this one has its exceptions.

The Coulomb gauge $\nabla \cdot \mathbf{A} = 0$ is not the only gauge condition popular in physics. Many physicists use the *Lorentz gauge*:

$$\nabla \cdot \mathbf{A} + \frac{\partial \psi}{c \partial t} = 0 \tag{1.394}$$

which causes (1.376) and (1.377) to have the forms

$$\nabla^2 \psi - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi = -4\pi\rho \tag{1.395}$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} = -\frac{4\pi}{c}\mathbf{j} \tag{1.396}$$

Now both the vector and scalar potentials obey the retarded wave equation. They combine to produce a four-vector which is invariant under a Lorentz transformation. The Lorentz invariance is very useful in many branches of physics, and the Lorentz gauge is used frequently. The Green's functions are not given for this gauge, but they are provided in books on field theory. Obviously both the scalar and vector parts are retarded.

Another gauge which is often used, but which does not have a formal name, is the condition that $\psi = 0$. The scalar potential is set equal to zero. In this case it is found that the

longitudinal vector potential is not zero but now plays an important role. In fact, the longitudinal part of the vector potential leads to an interaction between charges which is just the instantaneous Coulomb interaction. When $\psi = 0$ the longitudinal part of the vector potential plays a role that is identical to that of the scalar potential in the Coulomb gauge.

The Hamiltonian (1.360) is written in a gauge which has the scalar potential acting instantaneously, so that the Coulomb interaction is unretarded. This form of the Hamiltonian is consistent with either gauge $\nabla \cdot \mathbf{A} = 0$ or $\psi = 0$. One gets a different Hamiltonian for other choices of gauge.

1.5.2. Lagrangian

So far it has been proved that the Coulomb gauge makes the vector potential transverse and that the scalar potential acts instantaneously. Next it is shown that the Hamiltonian has the form indicated in Eq. (1.360), which is done by starting from the following Lagrangian (Schiff, 1968):

$$L = \frac{1}{2}\sum_i m\mathbf{v}_i^2 + \int \frac{d^3r}{8\pi}[\mathbf{E}(\mathbf{r})^2 - \mathbf{B}(\mathbf{r})^2] - \sum_i e_i\left[\psi(\mathbf{r}_i) - \frac{1}{c}\mathbf{v}_i \cdot \mathbf{A}(\mathbf{r}_i)\right]$$

This Lagrangian was chosen to produce Maxwell's equations as well as the classical equations of motion for a particle at $\mathbf{r}_i$, with charge e_i and velocity $\mathbf{v}_i$, in a magnetic and electric field. There are three variables η which are ψ, $\mathbf{A}$, $\mathbf{r}_i$. In terms of these variables the electric and magnetic fields are: $\mathbf{E} = -\mathbf{A}/c - \nabla\psi$, $\mathbf{B} = \nabla \times \mathbf{A}$. Each variable η is used to generate an equation from

$$\frac{d}{dt}\frac{\delta L}{\delta \dot{\eta}} + \sum_{\alpha=1}^{3}\frac{d}{dr_\alpha}\frac{\delta L}{\delta(\delta\eta/\delta r_\alpha)} - \frac{\delta L}{\delta \eta} = 0 \tag{1.397}$$

where the conjugate momentum is

$$P_\eta = \frac{\delta L}{\delta \dot{\eta}} \tag{1.398}$$

When the scalar potential ψ is chosen as the variable η,

$$\frac{\delta L}{\delta \psi} = -\rho(\mathbf{r}) = -\sum_i e_i\delta(\mathbf{r} - \mathbf{r}_i) \tag{1.399}$$

$$\frac{\partial L}{\partial(\partial\psi/\partial x)} = -\frac{E_x}{4\pi} \tag{1.400}$$

$$P_\psi = \frac{\delta L}{\delta \dot{\psi}} = 0 \tag{1.401}$$

This choice produces the Maxwell equation (1.366) from (1.397):

$$\nabla \cdot \mathbf{E} = 4\pi\rho \tag{1.402}$$

When one component of the scalar potential such as A_x is chosen as the variable in the Lagrangian,

$$c\frac{\delta L}{\delta A_x} = j_x = \sum_i e_i v_{ix}\delta(\mathbf{r} - \mathbf{r}_i) \tag{1.403}$$

$$\frac{\delta L}{\delta A_x} = -\frac{E_x}{4\pi} \tag{1.404}$$

$$\frac{\delta L}{\delta(\partial A_x/\partial y)} = -\frac{B_z}{4\pi} \tag{1.405}$$

When this equation is used in Lagrange's equation (1.397), the x-component of the Maxwell equation (1.367) is

$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c}\mathbf{j} \tag{1.406}$$

Choosing either A_y or A_z as the active variable will generate the y- and z- components of this equation. The Lagrangian does generate the two Maxwell equations which depend on particle properties.

Another important feature is that the momentum conjugate to the scalar potential is zero

$$P_\psi = 0 \tag{1.407}$$

The momentum variable conjugate to the vector potential is

$$\mathbf{P}_A = -\frac{\mathbf{E}}{4\pi c} \tag{1.408}$$

which is just proportional to the electric field. This relationship is important, since the quantization of the fields will require that the vector potential no longer commute with the electric field, since they are conjugate variables.

The other equations generated by this Lagrangian are the equations for particle motion in electric and magnetic fields. Here the variable in the Lagrangian is the particle coordinate r_α:

$$\frac{\delta L}{\delta r_{i\alpha}} = -e_i\left[\nabla_\alpha \psi(\mathbf{r}_i) - \frac{v_{i\delta}}{c}\nabla_\alpha A_\delta(\mathbf{r}_i)\right] \tag{1.409}$$

$$\frac{\delta L}{\delta v_{i\alpha}} = p_{i\alpha} = mv_{i\alpha} + \frac{e_i}{c}A_\alpha(\mathbf{r}_i) \tag{1.410}$$

In the last term of (1.409), repeated indices over δ imply summation. The equation deduced from Lagrange's equation is

$$0 = \frac{d}{dt}\left[m\mathbf{v}_i + \frac{e_i}{c}\mathbf{A}(\mathbf{r}_i)\right] + e_i\nabla\left[\psi - \frac{1}{c}\mathbf{v}\cdot\mathbf{A}(\mathbf{r})\right]$$

The total time derivative on the position dependence of the vector potential is interpreted as a hydrodynamic derivative:

$$\frac{d}{dt}\mathbf{A} = \frac{\partial}{\partial t}\mathbf{A} + \mathbf{v}_i\cdot\nabla\mathbf{A}$$

Then the equation may be rearranged into

$$m\frac{d}{dt}\mathbf{v}_i = e_i\left\{\mathbf{E}(\mathbf{r}_i) + \frac{1}{c}[\nabla(\mathbf{v}\cdot\mathbf{A}) - \mathbf{v}\cdot\nabla\mathbf{A}]\right\} = e_i\left[\mathbf{E}(\mathbf{r}_i) + \frac{1}{c}\mathbf{v}_i \times \mathbf{B}\right]$$

The last term is just the Lorentz force on a particle in a magnetic field since

$$\mathbf{v}\times\mathbf{B} = \mathbf{v}\times(\nabla\times\mathbf{A}) = \nabla(\mathbf{v}\cdot\mathbf{A}) - (\mathbf{v}\cdot\nabla)\mathbf{A} \tag{1.411}$$

The equation for $m\dot{\mathbf{v}}$ is Newton's law for a spinless particle in an electric and magnetic field. The Lagrangian is a suitable starting point for the quantization of the interacting system of particles and fields.

1.5.3. Hamiltonian

The Hamiltonian is derived from

$$H = \sum_i \dot{\eta}_i p_\eta - L \tag{1.412}$$

where the first summation is over all of the variables and their conjugate moments. In the present problem this summation includes the vector potential and the particle momentum; the scalar potential has no momentum. The Hamiltonian has the form

$$\begin{aligned} H = &\frac{1}{m}\sum_i \mathbf{p}_i\left[\mathbf{p}_i - \frac{e_i}{c}\mathbf{A}(\mathbf{r}_i)\right] + \int d^3r\left[\frac{1}{4\pi c}\dot{\mathbf{A}}\cdot\left(\frac{1}{c}\dot{\mathbf{A}} + \nabla\psi\right) - \frac{E^2 - B^2}{8\pi}\right] \\ &+ \sum_i e_i\psi(\mathbf{r}_i) - \frac{1}{2m}\sum_i\left[\mathbf{p}_i - \frac{e_i}{c}\mathbf{A}\right]^2 - \sum_i \frac{e_i}{mc}\mathbf{A}\cdot\left[\mathbf{p}_i - \frac{e_i}{c}\mathbf{A}\right] \end{aligned} \tag{1.413}$$

It may be collected into the form

$$\begin{aligned} H = &\frac{1}{2m}\sum_i\left[\mathbf{p}_i - \frac{e_i}{c}\mathbf{A}(\mathbf{r}_i)\right]^2 + \sum_i e_i\psi(\mathbf{r}_i) \\ &+ \int\frac{d^3r}{8\pi}\left[E^2 + B^2 - 2\nabla\psi\cdot\left(\frac{1}{c}\dot{\mathbf{A}} + \nabla\psi\right)\right] \end{aligned} \tag{1.414}$$

Terms which are the cross product between the scalar and vector potential parts of the electric field always vanish after an integration by parts,

$$\int d^3r(\nabla\psi\cdot\mathbf{A}) = -\int d^3r\psi(\nabla\cdot\mathbf{A}) = 0 \tag{1.415}$$

since the Coulomb gauge is used, wherein $\nabla\cdot\mathbf{A} = 0$. The terms involving the square of the scalar potential may also be reduced to an instantaneous interaction between charges using (1.381):

$$\int\frac{d^3r}{4\pi}\nabla\psi\cdot\nabla\psi = -\int\frac{d^3r}{4\pi}\psi\nabla^2\psi = \int d^3r\psi(\mathbf{r})\rho(\mathbf{r}) = \sum_i e_i\psi(\mathbf{r}_i) = \sum_{ij}\frac{e_ie_j}{r_{ij}} \tag{1.416}$$

With these simplifications, the Hamiltonian is written as

$$H = \frac{1}{2m}\sum_i \left[\mathbf{p}_i - \frac{e_i}{c}\mathbf{A}(\mathbf{r}_i)\right]^2 + \int \frac{d^3r}{8\pi}[E^2 + B^2] \tag{1.417}$$

The last terms are just the energy density of the electromagnetic fields. The first terms arise from the charged particles. This form of H is simple and instructive. However, our interest is in obtaining a quantized version. Write the electric field energy density as the separate parts of vector and scalar potential,

$$\begin{aligned}\int \frac{d^3r}{\pi}E^2 &= \int \frac{d^3r}{8\pi}\left[\frac{1}{c^2}\dot{\mathbf{A}}^2 + (\nabla\psi)^2 + \frac{2}{c}\nabla\psi\cdot\dot{\mathbf{A}}\right] \\ &= \frac{1}{2}\sum_{ij}\frac{e_ie_j}{r_{ij}} + \int \frac{d^3r}{8\pi c^2}\dot{\mathbf{A}}^2\end{aligned} \tag{1.418}$$

and the scalar potential just gives the interaction between charges.

The electric field is the conjugate momentum density of the vector potential. These two field variables must obey the following equal time commutation relation:

$$\left[A_\alpha(\mathbf{r}, t), -\frac{1}{4\pi}E_\beta(\mathbf{r}', t)\right] = i\delta_{\alpha\beta}\delta(\mathbf{r} - \mathbf{r}') \tag{1.419}$$

This commutation relation is satisfied by defining the vector potential in terms of creation and destruction operators in the form

$$A_\alpha(\mathbf{r}, t) = \sum_{\mathbf{k}\lambda}\left(\frac{2\pi\hbar c^2}{\nu\omega_\mathbf{k}}\right)^{1/2}\xi_\alpha(\mathbf{k}, \lambda)e^{i\mathbf{k}\cdot\mathbf{r}}(a_{\mathbf{k}\lambda}e^{-i\omega_\mathbf{k}t} + a^\dagger_{-\mathbf{k}\lambda}e^{-i\omega_\mathbf{k}t}) \tag{1.420}$$

There are two transverse modes, one for each transverse direction, and λ is the summation over these two modes. The unit polarization vector ξ_α gives the direction for each mode. The operators obey the commutation relations

$$[a_{\mathbf{k}\lambda}, a^\dagger_{\mathbf{k}'\lambda'}] = \delta_{\mathbf{k}\mathbf{k}'}\delta_{\lambda\lambda'} \tag{1.421}$$

$$[a_{\mathbf{k}\lambda}, a_{\mathbf{k}'\lambda'}] = 0 \tag{1.422}$$

The time derivative of the vector potential is

$$\dot{A}_\alpha(\mathbf{r}, t) = -i\sum_{\mathbf{k}\lambda}\left(\frac{2\pi\hbar c^2\omega_\mathbf{k}}{\nu}\right)^{1/2}\xi_\alpha(\mathbf{k}, \lambda)e^{i\mathbf{k}\cdot\mathbf{r}}(a_{\mathbf{k}\lambda}e^{-i\omega_\mathbf{k}t} - a^\dagger_{-\mathbf{k}\lambda}e^{-i\omega_\mathbf{k}t})$$

The electric field is

$$E_\alpha(\mathbf{r}, t) = -\frac{1}{c}\dot{A}_\alpha - \nabla_\alpha\psi \tag{1.423}$$

The scalar potential is not expressed in terms of operators, since its conjugate momentum is zero. It does not influence the commutator in (1.419). The commutation relation (1.419) is the commutator of the vector potential and its time derivative:

$$[A_\alpha(\mathbf{r}, t), \dot{A}_\beta(\mathbf{r}', t)] = 4\pi c i\delta_{\alpha\beta}\delta(\mathbf{r} - \mathbf{r}') \tag{1.424}$$

The vector potential in (1.420) has been chosen to give just this result:

$$[A_\alpha(\mathbf{r},t), \dot{A}_\beta(\mathbf{r}',t)] = i \sum_{\mathbf{k}\lambda\mathbf{k}'\lambda'} \left(\frac{2\pi\hbar c^2\omega_\mathbf{k}}{\nu}\right)^{1/2} \left(\frac{2\pi\hbar c^2}{\nu\omega'_\mathbf{k}}\right)^{1/2}$$
$$\times\, \xi_\alpha(\mathbf{k},\lambda)\xi_\beta(\mathbf{k},'\lambda')e^{i(\mathbf{k}\cdot\mathbf{r}+\mathbf{k}'\cdot\mathbf{r}')}2\delta_{\mathbf{k},-\mathbf{k}'}\delta_{\lambda\lambda'}$$
$$= 4\pi i c\delta_{\alpha\beta}\delta(\mathbf{r}-\mathbf{r}') \tag{1.425}$$

The various factors of 2π and $\omega_\mathbf{k}$ which enter (1.420) are selected so that the commutation relation (1.419) is satisfied for the usual commutation relations for the operators $a_{\mathbf{k}\lambda}$ and $a^\dagger_{\mathbf{k}\lambda}$. The expressions for the energy density of electric and magnetic fields have the form

$$\int d^3r \frac{1}{8\pi c^2}\dot{\mathbf{A}}^2(\mathbf{r},t) = -\sum_{\mathbf{k}\lambda}\frac{\hbar\omega_\mathbf{k}}{4}(a_{\mathbf{k}\lambda}a_{-\mathbf{k}\lambda}e^{-2it\omega_\mathbf{k}} + a^\dagger_{\mathbf{k}\lambda}a^\dagger_{-\mathbf{k}\lambda}e^{-2it\omega_\mathbf{k}}$$
$$- a_{\mathbf{k}\lambda}a^\dagger_{\mathbf{k}\lambda} - a^\dagger_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda}) \tag{1.426}$$

$$\int d^3r \frac{1}{8\pi}(\nabla\times\mathbf{A})^2 = \sum_{\mathbf{k}\lambda}\frac{\hbar c^2}{4\omega_\mathbf{k}}(\mathbf{k}\times\xi)^2(a_{\mathbf{k}\lambda}a_{-\mathbf{k}\lambda}e^{-2it\omega_\mathbf{k}} + a^\dagger_{\mathbf{k}\lambda}a^\dagger_{-\mathbf{k}\lambda}e^{-2it\omega_\mathbf{k}}$$
$$+ a_{\mathbf{k}\lambda}a^\dagger_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda}) \tag{1.427}$$

In the second term, the photon energy is $\omega_\mathbf{k} = ck$. This term may be added to the first, and the aa and $a^\dagger a^\dagger$ terms both cancel. Combining the results from (1.417) and (1.418),

$$H = \sum_i \frac{1}{2m}\left[\mathbf{p}_i - \frac{e_i}{c}\mathbf{A}(\mathbf{r}_i)\right]^2 + \frac{1}{2}\sum_{i\neq j}\frac{e_ie_j}{|\mathbf{r}_i-\mathbf{r}_j|} + \sum_{\mathbf{k}\lambda}\omega_\mathbf{k}a^\dagger_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda}$$

This form of the Hamiltonian is just the result which was asserted in the beginning in (1.360). The vector potential in (1.420) is expressed in terms of the creation and destruction operators of the photon field. The photon states behave as bosons: as independent harmonic oscillators. Each photon state of wave vector $\mathbf{k}$ and polarization λ has eigenstates pf the form

$$|n_{\mathbf{k}\lambda}\rangle = \frac{(a^\dagger_{\mathbf{k}\lambda})^{n_{\mathbf{k}\lambda}}}{(n_{\mathbf{k}\lambda}!)^{1/2}}|0\rangle \tag{1.428}$$

where $n_{\mathbf{k}\lambda}$ is an integer which is the number of photons in that state. The state $|0\rangle$ is the photon vacuum. The total energy in the free photon part of the Hamiltonian is

$$H_0|n_{k_1,\lambda_1}n_{k_2\lambda_2}\ldots n_{k_n\lambda_n}\rangle = \sum_{\mathbf{k}\lambda}\hbar\omega_\mathbf{k}(n_{\mathbf{k}\lambda}+\tfrac{1}{2})|n_{k_1,\lambda_1}n_{k_2\lambda_2}\ldots n_{k_n\lambda_n}\rangle \tag{1.429}$$

The Hamiltonian (1.360) has been derived for spinless, nonrelativistic particles. Certainly the most important relativistic term is the spin orbit interaction. The effects of spin also enter through the direct interaction of the magnetic moment with an external magnetic field.

1.6. PAIR DISTRIBUTION FUNCTION

In crystalline solids, the atoms are arranged in a regular array. If $\mathbf{R}_j$ denotes the position of the atoms, the summation of $\exp(i\mathbf{q}\cdot\mathbf{R}_j)$ over all the atoms yields zero unless $\mathbf{q}$ has particular values. These values are $\mathbf{q} = 0$ or else one of the reciprocal lattice vectors $\mathbf{G}$ of the solid. In either case, the factor $\exp(i\mathbf{G}\cdot\mathbf{R}_j) = 1$ so that the summation yields the number of atoms in the solid, which is N. This result is written as

$$\sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} = N\sum_{\mathbf{G}} \delta_{\mathbf{q}+\mathbf{G}} \tag{1.430}$$

where $\mathbf{q} = 0$ is in the term with $\mathbf{G} = 0$. Many materials are not crystalline, e.g., liquids, gases, and disordered solids. In discussing these quantities similar summations over particle locations are encountered. These summations still equal N for $\mathbf{q} = 0$, but they must be evaluated for $\mathbf{q} \neq 0$. This evaluation is done in terms of a function $S(\mathbf{q})$, which is called the *static structure factor* or the *static form factor.*

Define the density operator for the atomic locations:

$$\rho(\mathbf{q}) = \sum_j \exp(i\mathbf{q}\cdot\mathbf{R}_j) \tag{1.431}$$

$\rho(\mathbf{q})$ is an operator because the $\mathbf{R}_j$ describe the instantaneous location of each particle. It is assumed the particles are classical objects, so there is no need to deal with the quantum statistics associated with fermions or bosons. The average of this operator is defined as

$$\langle\rho(\mathbf{q})\rangle = \left\langle \sum_j \exp(i\mathbf{q}\cdot\mathbf{R}_j)\right\rangle \tag{1.432}$$

where the average is taken over the various configurations of the atoms. Some of these averages are defined here, without explaining how to find them from first principles. The latter is done in standard references (Hill, 1956; Percus, 1964).

The first average is over a single density operator:

$$\langle\rho(\mathbf{q})\rangle = N\delta_{\mathbf{q}=0} \tag{1.433}$$

$\langle\rho\rangle$ is zero for nonzero $\mathbf{q}$ because there is no restriction on the location of each atom. That is, in the averaging, each $\mathbf{R}_j$ can be anywhere in the material with equal likelihood since edge effects are ignored. The averaging is equivalent to the integral over a continuous distribution of values:

$$\langle\rho(\mathbf{q})\rangle = \frac{N}{\nu}\int d^3r e^{i\mathbf{q}\cdot\mathbf{r}} = N\delta_{\mathbf{q}=0} \tag{1.434}$$

The integral on the right just gives a delta function at $\mathbf{q} = 0$, which is the same result as (1.433).

The second interesting average is over a product of two density operators:

$$\langle\rho(\mathbf{q})\rho(\mathbf{q}')\rangle = \sum_{ij}\langle\exp(i\mathbf{q}\cdot\mathbf{R}_i)\exp(i\mathbf{q}'\cdot\mathbf{R}_j)\rangle$$

$$\langle\rho(\mathbf{q})\rho(\mathbf{q}')\rangle = N^2\delta_{\mathbf{q}=0}\delta_{\mathbf{q}'=0} + N\delta_{\mathbf{q}+\mathbf{q}'=0}S(\mathbf{q}) \tag{1.435}$$

Of course $\langle\rho(\mathbf{q})\rho(\mathbf{q}')\rangle = N^2$ when both $\mathbf{q}$ and $\mathbf{q}'$ are zero. If both $\mathbf{q}$ and $\mathbf{q}'$ are nonzero, the average vanishes unless $\mathbf{q}+\mathbf{q}'=0$. To prove this result, go to a center-of-mass coordinate system:

$$\mathbf{Q} = \mathbf{q}+\mathbf{q}', \qquad \mathbf{q} = \mathbf{k}+\frac{\mathbf{Q}}{2} \tag{1.436}$$

$$\mathbf{k} = \tfrac{1}{2}(\mathbf{q}-\mathbf{q}'), \qquad \mathbf{q}' = -\mathbf{k}+\frac{\mathbf{Q}}{2} \tag{1.437}$$

where the average is

$$\langle\rho(\mathbf{q})\rho(\mathbf{q}')\rangle = \sum_{ij}\left\langle \exp\left\{i\left[\mathbf{Q}\cdot\frac{\mathbf{R}_i+\mathbf{R}_j}{2}+\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)\right]\right\}\right\rangle \tag{1.438}$$

There is a strong correlation between the difference $\mathbf{R}_i-\mathbf{R}_j$ of two particle locations, but there is no correlation for the center of mass $\mathbf{R}_i+\mathbf{R}_j$. The term $\exp[i\mathbf{Q}\cdot(\mathbf{R}_i+\mathbf{R}_j)/2]$ averages to zero unless $\mathbf{Q}=0$, which gives $\mathbf{q}=-\mathbf{q}'=\mathbf{k}$. Whenever $\mathbf{q}\neq 0$ the average is

$$\langle\rho(\mathbf{q})\rho(\mathbf{q}')\rangle = \delta_{\mathbf{q}+\mathbf{q}'=0}\sum_{ij}\langle\exp[i\mathbf{q}\cdot(\mathbf{R}_i-\mathbf{R}_j)]\rangle \tag{1.439}$$

There is strong correlation between the relative positions of two atoms at short distances. If one atom is at a spot, the other will not be at the same spot because atoms repulse each other at short separations. The static structure factor $S(\mathbf{q})$ is defined as

$$N\delta_{\mathbf{q}=0}+S(\mathbf{q}) = \frac{1}{N}\sum_{ij}\langle\exp[i\mathbf{q}\cdot(\mathbf{R}_i-\mathbf{R}_j)]\rangle \tag{1.440}$$

$$= \frac{1}{N}\sum_{i}\left(1+\sum_{j\neq i}\exp[i\mathbf{q}\cdot(\mathbf{R}_i-\mathbf{R}_j)]\right) \tag{1.441}$$

One subtracts the term $N\delta_{\mathbf{q}=0}$. The right-hand side contains a summation over all particles at $\mathbf{R}_i$ and then a summation over all its neighbors $\mathbf{R}_i-\mathbf{R}_j$. The first term has $i=j$ and is independent of $\mathbf{q}$. When this formula is averaged over all possible positions of the impurities, it should give the same result for each particle $\mathbf{R}_i$. The summation over $\mathbf{R}_i$ just gives N:

$$N\delta_{\mathbf{q}=0}+S(\mathbf{q}) = 1+\left\langle\sum_{\Delta\mathbf{R}}{}' e^{i\mathbf{q}\cdot\Delta\mathbf{R}}\right\rangle \tag{1.442}$$

The prime on the summation means that $\Delta\mathbf{R}=0$ is excluded from the summation. The crucial average is the second term on the right. It averages over the relative positions $\Delta\mathbf{R}$ of the atoms near to the one at $\mathbf{R}_i$. This average can also be expressed as a function of $\mathbf{r}$. Define $g(\mathbf{r})$ as the *pair distribution function*. It is defined as the probability that another particle is at $\mathbf{r}$ if there is already one at $\mathbf{r}=0$. The normalization is chosen to give $g(\mathbf{r})\to 1$ at large distances. In terms of the pair distribution function, $S(\mathbf{q})$ may be written as

$$N\delta_{\mathbf{q}=0}+S(\mathbf{q}) = 1+n\int d^3r g(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.443}$$

where $n=N/\nu$ is the average density of the system. Also write

$$N\delta_{\mathbf{q}=0} = n\int d^3r e^{i\mathbf{q}\cdot\mathbf{r}} \tag{1.444}$$

and transpose this expression to the right to give the equivalent equations

$$S(\mathbf{q}) - 1 = n \int d^3 r e^{i\mathbf{q}\cdot\mathbf{r}}[g(\mathbf{r}) - 1]$$

$$g(\mathbf{r}) - 1 = \frac{1}{n} \int \frac{d^3 q}{(2\pi)^3} e^{-i\mathbf{q}\cdot\mathbf{r}}[S(\mathbf{q}) - 1] \tag{1.445}$$

The inverse Fourier transform may now be taken to give $g(\mathbf{r})$ as a function of $S(\mathbf{q})$. Both $g(\mathbf{r})$ and $S(\mathbf{q})$ go to unity at large values of their arguments. The functions $[g(\mathbf{r}) - 1]$ and $[S(\mathbf{q}) - 1]$ are both quantities which are short-ranged in their arguments. Their mutual Fourier transforms are well defined.

The function $S(\mathbf{q})$ is determined experimentally by scattering from targets. Afterwards, $g(\mathbf{r})$ is found by a numerical Fourier transform. In liquids they depend only upon the magnitude of their vector arguments: $g(\mathbf{r})$ and $S(\mathbf{q})$. The well-known results for liquid ^{4}He are shown in Fig. 1.3, which is from Mozer *et al.* (1974). The pair distribution function $g(r)$ should vanish according to the dashed line at short distances because the atoms do not penetrate each other at the low thermal energies of the liquid. The numerical transform shown by the solid line has a spurious peak for $R < 2$ Å because of inaccuracies in $S(k)$ for $k \to \infty$.

The higher moments such as $\langle \rho(\mathbf{q})\rho(\mathbf{q}')\rho(\mathbf{q}'')\rangle$ would also be interesting were they known. But they are not easily obtainable from experiments nor from theory.

In the model of the homogeneous electron gas, the electrons are not ordered. In this case one can also describe relative behavior of pairs of particles by a pair distribution function $g(r)$. This quantity is always defined as the average over the motion of the particles. In a quantum system the diagonal density matrix is defined as the square of the N-particle wave function:

$$\rho_N(\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N) = |\Psi(\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N)|^2 \tag{1.446}$$

It is normalized so that unity is obtained when integrating over all the coordinates:

$$1 = \int d^3 r_1 d^3 r_2 \cdots d^3 r_N \rho_N(\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N) \tag{1.447}$$

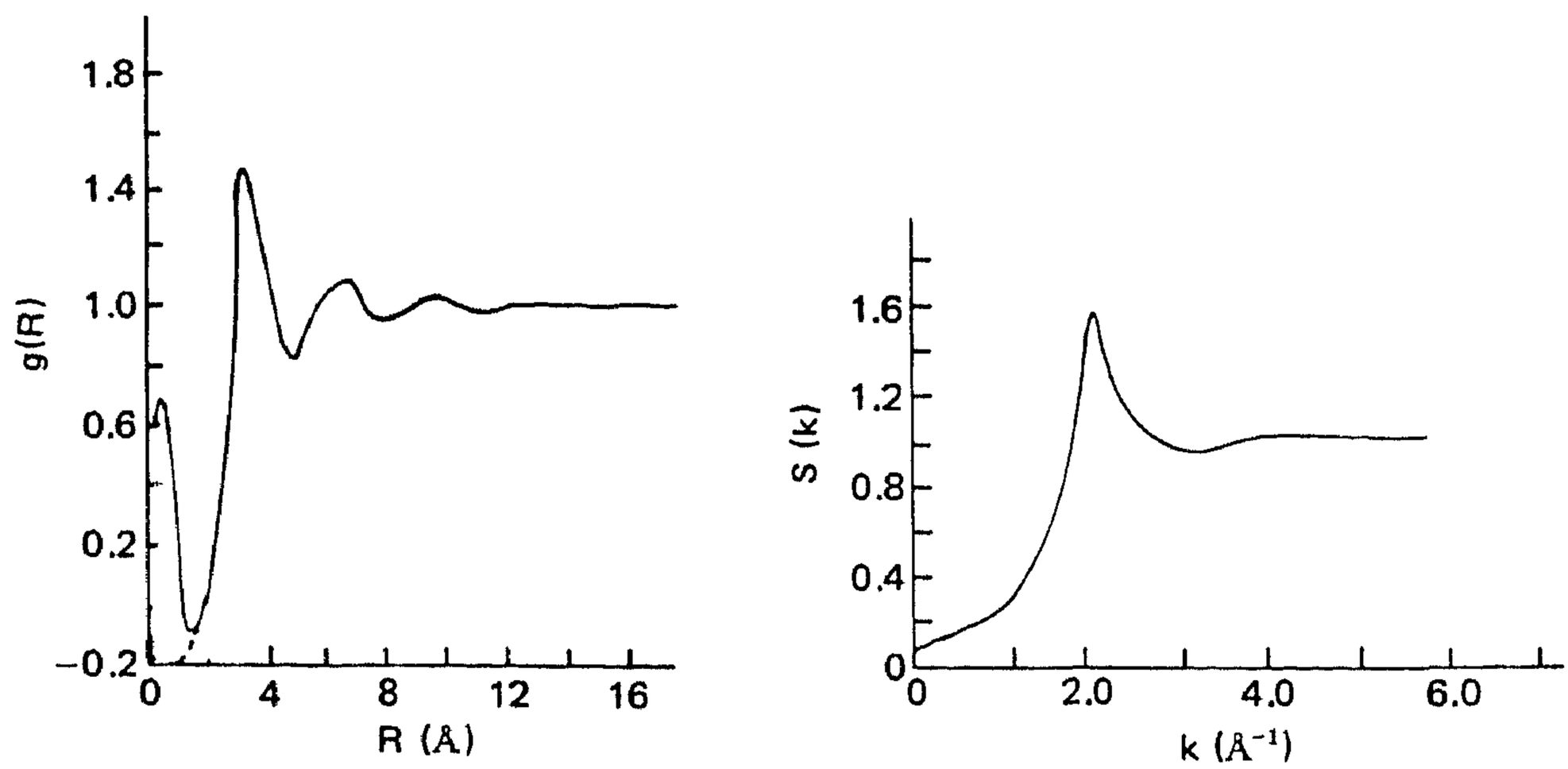


FIGURE 1.3 $g(R)$ (left) and $S(k)$ (right) for liquid ^{4}He. $S(k)$ is obtained by X-ray scattering measurements, and $g(R)$ is found by the numerical transform. Uncertainties in $S(k)$ at large k produce the extraneous structure in $g(R)$ at low R, which is to be ignored. The dashed line shows the proper extrapolation. $\rho = 0.1628\,\text{g/cm}^3$, $T = 2.86\,\text{K}$. Source: Mozer *et al.* (1974) (used with permission).

Since the many-particle wave function Ψ is either even or odd under the interchange of two identical particles, the density matrix is unchanged by this operator. Any two particles can be called $\mathbf{r}_1$ and $\mathbf{r}_2$. Their relative behavior, after averaging over all the other particles, is obtained by integrating over the coordinates of all the other particles:

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \int d^3r_3 d^3r_4 \cdots d^3r_N \rho_N(\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N) \tag{1.448}$$

$\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ is the average two-particle density matrix. The "integration" over other particles included spatial integration, spin averages, etc. Similarly, the one-particle density matrix is that obtained by integrating over all but one of the particles:

$$\rho_1(\mathbf{r}_1) = \int d^3r_2 d^3r_3 \cdots d^3r_N \rho_N(\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N) \tag{1.449}$$

$$= \int d^3r_2 \rho_2(\mathbf{r}_1, \mathbf{r}_2) \tag{1.450}$$

$$1 = \int d^3r_1 \rho(\mathbf{r}_1) \tag{1.451}$$

For homogeneous systems, the two-particle density matrix $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ can only depend on the relative positions $\mathbf{r}_1 - \mathbf{r}_2$ of the two coordinates. This quantity must be related to the pair distribution function. They are in fact proportional. The constant of proportionality is determined from the condition that $g(\mathbf{r}) \to 1$ at large separation, so

$$N(N-1)\rho_2(\mathbf{r}_1, \mathbf{r}_2) = n^2 g(\mathbf{r}_1 - \mathbf{r}_2) \tag{1.452}$$

As an example of particular behavior, evaluate $g(\mathbf{r})$ for electrons using the Hartree–Fock approximation. The many-particle wave function is just an N-dimensional Slater determinant, so the N-particle density matrix is just the square of this determinant:

$$\rho_N(\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N) = \frac{1}{N!} \begin{vmatrix} \psi_{\lambda_1}(\mathbf{r}_1) & \psi_{\lambda_2}(\mathbf{r}_1) & \cdots & \psi_{\lambda_N}(\mathbf{r}_1) \\ \psi_{\lambda_1}(\mathbf{r}_2) & \psi_{\lambda_2}(\mathbf{r}_2) & \cdots & \psi_{\lambda_N}(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\lambda_1}(\mathbf{r}_N) & \psi_{\lambda_2}(\mathbf{r}_N) & \cdots & \psi_{\lambda_N}(\mathbf{r}_N) \end{vmatrix}^2 \tag{1.453}$$

The single-particle orbitals $\psi_{\lambda_1}(\mathbf{r}_1)$ are assumed to be orthogonal for different states λ. A theorem states (see Parr, 1964) that if one integrates over all but two coordinates, the two-particle density matrix, from this Slater determinant, is the sum of all pairs of two-particle wave functions:

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{N(N-1)} \sum_{\lambda_1 \neq \lambda_2} \frac{1}{2!} \begin{vmatrix} \psi_{\lambda_1}(\mathbf{r}_1) & \psi_{\lambda_2}(\mathbf{r}_1) \\ \psi_{\lambda_1}(\mathbf{r}_2) & \psi_{\lambda_2}(\mathbf{r}_2) \end{vmatrix}^2 \tag{1.454}$$

The summation is taken over all pairs of occupied states. The two-dimensional Slater determinant can be expanded to give the following equivalent result:

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2N(N-1)} \sum_{\lambda_1 \neq \lambda_2} [\psi_{\lambda_1}(\mathbf{r}_1)\psi_{\lambda_2}(\mathbf{r}_2) - \psi_{\lambda_1}(\mathbf{r}_2)\psi_{\lambda_2}(\mathbf{r}_1)]^2 \tag{1.455}$$

A further integration gives the one-particle density matrix as the summation over all occupied states of the square of the wave function:

$$\rho_1(\mathbf{r}_1) = \frac{1}{2N(N-1)} \sum_{\lambda_1 \neq \lambda_2} [|\psi_{\lambda_1}(\mathbf{r})|^2 + |\psi_{\lambda_2}(\mathbf{r})|^2] \tag{1.456}$$

$$\rho_1(\mathbf{r}) = \frac{1}{N} \sum_{\lambda_1} |\psi_{\lambda_1}(\mathbf{r})|^2 \tag{1.457}$$

These theorems apply to any Hartree–Fock system with any orbitals. For homogeneous systems, it is assumed the wave functions are translationally invariant. If they are plane waves with spin up α, the two-particle density matrix is

$$\psi = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{\mathrm{v}}}\alpha \tag{1.458}$$

$$\rho_2(\mathbf{r}, \mathbf{r}') = \frac{1}{\mathrm{v}^2 N(N-1)} \sum_{\mathbf{k}\mathbf{k}'} [1 - e^{i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}-\mathbf{r}')}]\alpha_1\alpha_2 \tag{1.459}$$

Using (1.452) gives a pair distribution function of the form

$$g(\mathbf{r} - \mathbf{r}') = \frac{1}{N^2} \sum_{\mathbf{k}\mathbf{k}'} [1 - e^{i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}-\mathbf{r}')}]\alpha_1\alpha_2 \tag{1.460}$$

In most systems, the electrons do not all have spin up, and one also has to average over all spin coordinates to get $g(\mathbf{r})$. This procedure is discussed in Chapter 5.

Problems

1. Solve the classical vibrational modes of a one-dimensional chain of atoms of type A and B. They alternate on the chain with masses m_A and m_B. The harmonic spring between atoms has spring constant K.

2. Write down the Hamiltonian of Problem 1. Solve it, and show that it may be reduced to the form

$$H = \sum_{k\lambda} \omega_{k\lambda}(a^\dagger_{k\lambda} a_{k\lambda} + \tfrac{1}{2}) \tag{1.461}$$

where the $\omega_{k\lambda}$ are the classical normal modes.

3. Find the exact solution to

$$H = E_0 a^\dagger a + E_1(a^\dagger a^\dagger + aa) \tag{1.462}$$

where E_0 and E_1 are constants and a and $a^\dagger$ are boson operators.

4. Solve the Hamiltonian below with a canonical transformation:

$$H = E_0 a^\dagger a + F(a + a^\dagger) \tag{1.463}$$

$$\bar{H} = e^s H e^{-s}, \qquad s = \lambda(a - a^\dagger) \tag{1.464}$$

a. Show that $\lambda = \lambda^*$ since H is Hermitian.
b. Use the expansion (1.356), and show that only a few terms in the series are finite—the remainder vanish.
c. Find the choice of λ which reduces H to (1.50).

5. Consider a fermion system which has three energy states with eigenvalues E_1, E_2, and E_3. There also exist matrix elements which connect these states and permit transitions between them: M_{12}, M_{23}, and M_{13}

(a) Write down the Hamiltonian for this system in terms of creation and destruction operators.
(b) Determine the eigenvalue equation for this system.

6. Consider a tight-binding solid which has alternate atoms of type A and B. The electron Hamiltonian in the nearest neighbor model has the form

$$= w \sum_{i\delta} (a_i^\dagger b_{i+\delta} + b_{i+\delta}^\dagger a_i) + \sum_i [B b_i^\dagger b_i + A a_i^\dagger a_i] \tag{1.465}$$

where a_i and b_j are electron operators for atoms of type A and B. Find the exact eigenvalues of this Hamiltonian.

7. Calculate the exact partition function

$$e^{-\beta\Omega} = Z = \mathrm{Tr}\, e^{-\beta(H-\mu N)} \tag{1.466}$$

for the Hubbard model (1.182) in the atomic limit. Then give the expression for the average number of electrons $\bar{N} = -(\partial\Omega/\partial\mu)$.

8. Take a three-atom chain with periodic boundary conditions. Each atom has one orbital state with two spin configurations.

(a) Solve the nearest neighbor tight-binding model for zero, one, two, or three electrons with spin up, and give the partition function. Show that it is the same as (1.176).
(b) Give the partition function for the XY model, and show that it is different.

9. Give the partition function for the Hubbard model for a two-atom chain with periodic boundary conditions.

10. What is the energy current operator $\mathbf{j}_E$ for the Hubbard model (1.182)?

11. Show that the energy current in the one-dimensional harmonic chain (1.50) is

$$j_E = \frac{1}{2} \sum_k \frac{d(\omega_k^2)}{dk} a_k^\dagger a_k \tag{1.467}$$

12. Assume that the Hamiltonian in the tight-binding model has additional terms describing how phonons $(b_i, b_i^\dagger)$ on a site i interact with electrons n_i when they are on the site. How do these terms affect the heat current?

$$\sum_i [\omega_0 b_i^\dagger b_i + M n_i (b_i^\dagger + b_i)] \tag{1.468}$$

13. Explicitly verify the equation of continuity,

$$\frac{\partial}{\partial t}\rho(\mathbf{r}, t) + \nabla \cdot \mathbf{j}(\mathbf{r}, t) = 0 \tag{1.469}$$

for a gas of free fermion particles which have

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} \tag{1.470}$$

$$\rho(\mathbf{r}, t) = \frac{1}{V} \sum_{\mathbf{k}\mathbf{q}\sigma} e^{-i\mathbf{q}\cdot\mathbf{r}} c^{\dagger}_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma} \tag{1.471}$$

$$\mathbf{j}(\mathbf{r}, t) = \frac{1}{mV} \sum_{\mathbf{k}\mathbf{q}\sigma} e^{-i\mathbf{q}\cdot\mathbf{r}} (\mathbf{k} + \tfrac{1}{2}\mathbf{q}) c^{\dagger}_{\mathbf{k}+\mathbf{q}\sigma} c_{\mathbf{k}\sigma} \tag{1.472}$$

14. Find the effective interaction $V_R(r)$ between two fixed electrons (1.247) using the deformation potential interaction (1.249) and a Debye spectrum.

15. Find the exact solution in one dimension and the partition function for the XY model with a magnetic field in the z direction.

16. The Holstein–Primakoff transformation between spin one-half operators and bosons is

$$S_l^{(z)} = \tfrac{1}{2} - a_l^{\dagger} a_l \tag{1.473}$$

$$S_l^{(+)} = (1 - a_l^{\dagger} a_l)^{1/2} a_l \tag{1.474}$$

$$S_l^{(-)} = a_l^{\dagger} (1 - a_l^{\dagger} a_l)^{1/2} \tag{1.475}$$

where the a_l are boson operators.

(a) Use these definitions to evaluate the following commutators:

$$[S_l^{(+)}, S_m^{(-)}]; \quad [S_l^{(z)}, S_m^{(-)}] \tag{1.476}$$

(b) Write out the Heisenberg Hamiltonian in terms of the a operators.

(c) At low temperature, where the number of excitations $\bar{n}_l \leqslant \langle a_l^{\dagger} a_l \rangle$ is small, one can simplify the Hamiltonian by neglecting all terms of the form $a^{\dagger} a^{\dagger} a a$ which describe the scattering of two excitations. Then one can solve the Hamiltonian exactly. What are the eigenvalues?

Chapter 2

Green's Functions at Zero Temperature

Many-body calculations are often done for model systems at zero temperature. Of course, real experimental systems are never at zero temperature, although they are often at low temperature. Many quantities are not very sensitive to temperature, particularly at low temperature. Zero temperature calculations are useful even for describing real systems. Furthermore, the zero temperature property of a system is an important conceptual quantity—the ground state of an interacting system. A system is often described as its ground state plus its excitations, and the ground state may be deduced from a zero temperature calculation. Many zero temperature calculations have been done to deduce, for example, the ground state of the homogeneous electron gas or the ground state of superfluid ^{4}He.

Some workers believe that zero temperature calculations are easier to perform. There are usually more terms at nonzero temperature, which is a nuisance. We do not completely share this viewpoint. The Matsubara methods, which are described in Chapter 3 for nonzero temperature, are very easy to use. We usually do zero temperature calculations by first finding the nonzero temperature formulas and then taking the limit of zero temperature. However, sometimes it is easier to do zero temperature calculations from the beginning. The zero temperature formalism is a necessary part of one's calculational machinery.

It is presumed that one is trying to solve a Hamiltonian which cannot be solved exactly. One does not need Green's functions if the problem may be solved exactly. Very few exact results were obtained by Green's functions which were not first obtained by conventional theoretical techniques. It is assumed that one is trying to deduce the properties of some system described by a Hamiltonian H which may not be solved exactly. The usual approach is to set

$$H = H_0 + V \tag{2.1}$$

where H_0 is a Hamiltonian which may be solved exactly. The term V represents all the remaining parts of H. One tries to choose H_0 so that the effects of V are small. The basic procedure is to start with a system completely described by H_0. The effects of V are introduced, and then calculations are done to find how V changes the properties. These steps are the basic procedure in many-body theory.

2.1. INTERACTION REPRESENTATION

Most readers of this book should be familiar with the interaction representation. A short discussion of this topic will be presented anyway, as a refresher. There are three representations which will be discussed: Schrödinger, Heisenberg, and interaction.

2.1.1. Schrödinger

Elementary quantum mechanics is taught in the *Schrödinger representation*, which is based on the formula ($\hbar = 1$)

$$i\frac{\partial}{\partial t}\psi(t) = H\psi(t) \tag{2.2}$$

which has the formal operator solution

$$\psi(t) = e^{-iHt}\psi(0) \tag{2.3}$$

The use of this formula requires some assumptions:

1. The wave functions $\psi(t)$ are time dependent, even if this dependence is only a simple factor of $\exp(-iEt)$.
2. Operators such as the Hamiltonian H are taken to be independent of time.

2.1.2. Heisenberg

It is possible to solve quantum mechanical problems another way which gives exactly the same answers yet uses methods that look quite different. The *Heisenberg representation* has the following properties:

1. The wave functions are independent of time.
2. The operators are time dependent, and this dependence is given by

$$O(t) = e^{iHt}O(0)e^{-iHt} \tag{2.4}$$

or, equivalently, one is trying to solve the equation of motion which is derived from a time derivative:

$$i\frac{\partial}{\partial t}O(t) = [O(t),\ H] \tag{2.5}$$

Note that the time dependences of these two representations appear to be contrary. Yet it was asserted they give the same answer. To prove this identity takes too much space, so the result is just made plausible.

In physics one is usually trying to evaluate matrix elements in order to determine transition rates. In the Schrödinger representation, the matrix element of the operator $O(0)$ between two states is

$$\langle\psi_1^\dagger(t)O(0)\psi_2(t)\rangle = \langle\psi^\dagger(0)e^{iHt}O(0)e^{-iHt}\psi_2(0)\rangle \tag{2.6}$$

In the Heisenberg representation one obtains the result

$$\langle\psi_1^\dagger(0)O(t)\psi_2(0)\rangle = \langle\psi^\dagger(0)e^{iHt}O(0)e^{-iHt}\psi_2(0)\rangle \tag{2.7}$$

The two representations produce the same result. In order to understand this assertion, first recall that the Hamiltonian H will include some interaction terms. In the Heisenberg representation, these interactions act upon the operators, thereby changing them. When taking matrix elements, these changed operators are projected back upon the unchanged states in order to see how much the operators are changed by the interactions. On the other hand, the Schrödinger picture leaves the operators fixed and instead has the interactions affect the wave functions. Then the new wave functions are used to evaluate the matrix elements of the unchanged operators. The result is the same either way.

2.1.3. Interaction

The *interaction representation* is another way of doing things. Here both the wave functions and the operators are time dependent. The Hamiltonian is separated into two parts,

$$H = H_0 + V \tag{2.8}$$

where H_0 is the unperturbed part, while the V are the interactions. At the moment the exact form of V is left unspecified, and we merely note that it may be either, some, or all of the interactions discussed in Chapter 1. Usually H_0 is selected as a Hamiltonian which is exactly solvable.

Operators and wave functions in the interaction representation will be denoted by a caret, such as in $\hat{V}$. This notation will distinguish their time dependence from that of the other representations:

1. Operators have a time dependence

$$\hat{O}(t) = e^{iH_0t} O e^{-iH_0t} \tag{2.9}$$

2. Wave functions have a time dependence

$$\hat{\psi}(t) = e^{iH_0t} e^{-iHt} \psi(0) \tag{2.10}$$

It is assumed that $[H_0, V] \neq 0$. If these operators do commute, the problem is usually too trivial to require many-body theory. When these operators do not commute, the exponentials cannot be combined, because

$$e^A e^B = e^{A+B} \tag{2.11}$$

only if $[A, B] = 0$.

Before going any further, let us check to see that this choice of time dependence does produce the same matrix elements as before:

$$\begin{aligned} \langle \hat{\psi}_1^\dagger(t) \hat{O}(t) \hat{\psi}_2(t) \rangle &= \langle \psi_1^\dagger(0) e^{iHt} e^{-iH_0t} (e^{iH_0t} O e^{-iH_0t}) e^{iH_0t} e^{-iHt} \psi_2(0) \rangle \\ &= \langle \psi^\dagger(0) e^{iHt} O(0) e^{-iHt} \psi_2(0) \rangle \end{aligned} \tag{2.12}$$

It does. The time dependence of the operators is governed by the unperturbed Hamiltonian. Next show that the time dependence of the wave functions is governed by the interactions

$$\begin{aligned}\frac{\partial}{\partial t}\hat{\psi}(t) &= ie^{iH_0t}(H_0 - H)e^{-iHt}\psi(0)\\ &= -ie^{iH_0t}Ve^{-iHt}\psi(0)\\ &= -ie^{iH_0t}Ve^{-iH_0t}[e^{iH_0t}e^{-iHt}\psi(0)]\\ \frac{\partial}{\partial t}\hat{\psi}(t) &= -i\hat{V}(t)\hat{\psi}(t)\end{aligned} \tag{2.13}$$

This result proves the assertion that the time dependence of $\hat{\psi}(t)$ is determined by $\hat{V}(t)$. An operator was introduced into Eq. (2.10) which is defined as $U(t)$:

$$U(t) = e^{iH_0t}e^{-iHt} \tag{2.14}$$

This function has the value of unity at $t = 0$: $U(0) = 1$. Furthermore, it obeys a differential equation which can be written in the interaction representation:

$$\begin{aligned}\frac{\partial}{\partial t}U(t) &= ie^{iH_0t}(H_0 - H)e^{-iHt}\\ &= -ie^{iH_0t}V(e^{-iH_0t}e^{iH_0t})e^{-iHt}\\ &= -i\hat{V}(t)U(t)\end{aligned} \tag{2.15}$$

In order to solve this equation, one way of proceeding is by integrating both sides of the equation with respect to time:

$$U(t) - U(0) = -i\int_0^t dt_1\hat{V}(t_1)U(t_1) \tag{2.16}$$

Rearranging gives ($U(0) = 1$)

$$U(t) = 1 - i\int_0^t dt_1\hat{V}(t_1)U(t_1) \tag{2.17}$$

If this equation is iterated repeatedly, then

$$\begin{aligned}U(t) &= 1 - i\int_0^t dt_1\hat{V}(t_1) + (-i)^2\int_0^t dt_1\int_0^{t_1} dt_2\hat{V}(t_1)\hat{V}(t_2) + \cdots\\ &= \sum_{n=0}^{\infty}(-i)^n\int_0^t dt_1\int_0^{t_1} dt_2\cdots\int_0^{t_{n-1}} dt_n\hat{V}(t_1)\hat{V}(t_2)\ldots\hat{V}(t_n)\end{aligned} \tag{2.18}$$

At this point it is convenient to introduce the time-ordering operator T, which should not be confused with the temperature. The T operator acts upon a group of time-dependent operators,

$$T[\hat{V}(t_1)\hat{V}(t_2)\hat{V}(t_3)] \tag{2.19}$$

and is just an instruction to arrange the operators with the earliest times to the right. For example,

$$T[\hat{V}(t_1)\hat{V}(t_2)\hat{V}(t_3)] = \hat{V}(t_3)\hat{V}(t_1)\hat{V}(t_2) \qquad \text{if } t_3 > t_1 > t_2 \tag{2.20}$$

It helps to introduce the following step function:

$$\begin{aligned}\Theta(x) &= 1 \qquad \text{if } x > 0\\ &= 0 \qquad \text{if } x < 0\\ &= \tfrac{1}{2} \qquad \text{if } x = 0\end{aligned} \tag{2.21}$$

For two operators, the explicit definition of T ordering gives

$$T[\hat{V}(t_1)\hat{V}(t_2)] = \Theta(t_1 - t_2)\hat{V}(t_1)\hat{V}(t_2) + \Theta(t_2 - t_1)\hat{V}(t_2)\hat{V}(t_1) \tag{2.22}$$

Of course, if $V(t_1)$ and $V(t_2)$ commuted with each other, then the order of the operators is unimportant. The T ordering needs to be applied only to operators which do not commute at different times. Now rearrange the integral by using the above identity:

$$\begin{aligned}&\frac{1}{2!}\int_0^t dt_1 \int_0^t dt_2 T[\hat{V}(t_1)\hat{V}(t_2)]\\ &= \frac{1}{2!}\int_0^t dt_1 \int_0^{t_1} dt_2 \hat{V}(t_1)\hat{V}(t_2) + \frac{1}{2!}\int_0^t dt_2 \int_0^{t_2} dt_1 \hat{V}(t_2)\hat{V}(t_1)\end{aligned} \tag{2.23}$$

The second term on the right-hand side is equal to the first, which is easy to see by just redefining the integration variables $t_1 \to t_2$, $t_2 \to t_1$. These steps give

$$\frac{1}{2!}\int_0^t dt_1 \int_0^t dt_2 T[\hat{V}(t_1)\hat{V}(t_2)] = \int_0^t dt_1 \int_0^{t_1} dt_2 \hat{V}(t_1)\hat{V}(t_2) \tag{2.24}$$

Similarly, one can show that

$$\begin{aligned}&\frac{1}{3!}\int_0^t dt_1 \int_0^t dt_2 \int_0^t dt_3 T[\hat{V}(t_1)\hat{V}(t_2)\hat{V}(t_3)]\\ &= \int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 \hat{V}(t_1)\hat{V}(t_2)\hat{V}(t_3)\end{aligned} \tag{2.25}$$

Returning to the expansion of $U(t)$ gives

$$U(t) = 1 + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int_0^t dt_1 \int_0^t dt_2 \cdots \int_0^t dt_n T[\hat{V}(t_1)\hat{V}(t_2)\cdots\hat{V}(t_n)] \tag{2.26}$$

This expansion may be abbreviated by writing it as

$$U(t) = T\exp\left[-i\int_0^t dt_1 \hat{V}(t_1)\right] \tag{2.27}$$

However, it should always be kept in mind that the exponential form is really just a shorthand for the series definition (2.26). The T-ordered series definition is really equivalent to the original expansion (2.18).

2.2. S MATRIX

It was shown in the previous section that the wave function in the interaction representation had a time dependence given by

$$\hat{\psi}(t) = U(t)\hat{\psi}(0) \tag{2.28}$$

Now define the S matrix as the operator $S(t, t')$ which changes the wave function $\hat{\psi}(t')$ into $\hat{\psi}(t)$:

$$\hat{\psi}(t) = S(t, t')\hat{\psi}(t') \tag{2.29}$$

From the original definition

$$\hat{\psi}(t) = U(t)\hat{\psi}(0) = S(t, t')U(t')\hat{\psi}(0) \tag{2.30}$$

which produces the result

$$S(t, t') = U(t)U^{\dagger}(t') \tag{2.31}$$

Now examine some properties of this operator. The first two of the following identities may be proved in a trivial way:

1. $S(t, t) = 1 = U(t)U^{\dagger}(t) = e^{iH_0t}e^{-iHt}(e^{iHt}e^{-iH_0t})$
2. $S^{\dagger}(t, t') = U(t')U^{\dagger}(t) = S(t', t)$
3. $S(t, t')S(t', t'') = S(t, t'')$

The third identity can be shown by appealing to the original definition:

$$\hat{\psi}(t) = S(t, t')\hat{\psi}(t') = S(t, t')S(t', t'')\hat{\psi}(t'') = S(t, t'')\hat{\psi}(t'')$$

Finally, it is shown that $S(t, t')$ can also be expressed as a time-ordered operator,

$$\frac{\partial}{\partial t}S(t, t') = \frac{\partial}{\partial t}U(t)U^{\dagger}(t') = -i\hat{V}(t)S(t, t') \tag{2.32}$$

which has the solution

$$S(t, t') = T\exp\left[-i\int_{t'}^{t} dt_1\,\hat{V}(t_1)\right] \tag{2.33}$$

The function $\hat{\psi}(0) \equiv \psi(0)$ was introduced in the discussion of the previous section. It is a wave function in the Heisenberg representation, so that it is independent of time. If a Schrödinger wave function is defined as

$$\psi_s(t) = e^{-iHt}\psi(0) \tag{2.34}$$

then $\psi(0)$ is also the Schrödinger wave function at $t = 0$. Define

$$\hat{\psi}(t) = U(t)\psi(0) \tag{2.35}$$

Then $\psi(0)$ is also the $t = 0$ wave function in the interaction representation. At zero temperature the only wave function of special interest is the ground state wave function. For Green's functions it is necessary to define $\psi(0)$ as the exact ground state wave function. Since the total Hamiltonian is H, the exact ground state must have the lowest eigenvalue of this Hamiltonian. An immediate problem is that the eigenvalues or eigenstates of the Hamiltonian

H are not known initially. In fact, that is exactly the kind of information we are trying to obtain by using Green's functions.

There is a problem in that the formalism is based on the wave function $\psi(0)$ which is not yet known. Now in the interaction representation set $H = H_0 + V$, where H_0 is chosen to be sufficiently simple that its eigenvalues and eigenstates are known. Let the lowest eigenvalue of H_0—its ground state—be denoted ϕ_0. Somehow the unknown wave function $\psi(0)$ must be determined in terms of the known wave function ϕ_0.

The relationship between the two ground states $\psi(0)$ and ϕ_0 at zero temperature was established by Gell-Mann and Low (1951):

$$\psi(0) = S(0, -\infty)\phi_0 \tag{2.36}$$

This result will not be proven: instead, it will be made plausible. One known result is that

$$\hat{\psi}(t) = S(t, 0)\psi(0) \tag{2.37}$$

Operate by $S(0, t)$ and get

$$\psi(0) = S(0, t)\hat{\psi}(t) \tag{2.38}$$

since $S(0, t)S(t, 0) = 1$ by the use of the previous theorems. Let $t \to -\infty$, and find

$$\psi(0) = S(0, -\infty)\hat{\psi}(-\infty) \tag{2.39}$$

The important assertion is that $\hat{\psi}(-\infty)$ is equal to ϕ_0. The traditional argument is that one starts in the dim past ($t \to -\infty$) with a wave function ϕ_0 which does not contain the effects of the interaction V. The operator $S(0, -\infty)$ brings this wave function adiabatically up to the present $t = 0$. It is a wave function which does contain the effects of the interaction V, so that it is an eigenstate of H.

There is an additional property of these states which is needed for the discussion of Green's functions. As $t \to +\infty$, then

$$\hat{\psi}(\infty) = S(\infty, 0)\psi(0) \tag{2.40}$$

One possible assumption is that $\hat{\psi}(\infty)$ must be related to ϕ_0. If they are equal except for a phase factor L:

$$\phi_0 e^{iL} = \hat{\psi}(\infty) = S(\infty, 0)\psi(0) = S(\infty, -\infty)\phi_0 \tag{2.41}$$

$$e^{iL} = \langle\phi_0|S(\infty, -\infty)|\phi_0\rangle \tag{2.42}$$

Alternatives to this assumption are discussed in Sec. 2.9.

2.3. GREEN'S FUNCTIONS

Most of this book is concerned with Green's functions for only three types of particles: electron, phonon, and photon. The electron and phonon cases will be discussed here. The photon case is badly muddled by the several available choices of electromagnetic gauge. Section 2.10 is devoted to discussing these complications, and the photon Green's functions will be treated there. But keep in mind that the basic proofs for the photon case are essentially identical to the phonon case.

At zero temperature the electron Green's function is defined as

$$G(\lambda, t - t') = -i\langle|TC_\lambda(t)C_\lambda^\dagger(t')|\rangle \tag{2.43}$$

The quantum number λ can be anything depending on the problem of interest. Quite often it shall be taken to be the quantum numbers of the free-electron gas $\lambda = (\mathbf{p}, \sigma)$, where $\mathbf{p}$ is the wave vector and σ is the spin. At zero temperature the state $|\rangle$ must be the ground state. If the Hamiltonian of the problem is chosen to be H, then $|\rangle$ is the ground state of H, and therefore it is an eigenstate of H.

Of course, initially the ground state or any other eigenstate of H is not known since that is what is to be determined by using Green's functions. Write $H = H_0 + V$, where H_0 is the unperturbed part, while V is the interaction. Choose H_0 so that its eigenstates are known. In Sec. 1.3 the C_λ are defined in terms of a complete set of states ψ_λ. Now select this complete set of states as the eigenstates of the unperturbed Hamiltonian H_0. In the definition of the Green's function the C_λ represent states of H_0, while the ground state $|\rangle$ is an eigenstate of H. Furthermore, (2.43) is defined in the Heisenberg representation, so that $|\rangle$ is independent of time, while $C_\lambda(t)$ is given by the usual result

$$C_\lambda(t) = e^{iHt}C_\lambda e^{-iHt} \tag{2.44}$$

One way of understanding the Green's function is to observe that it describes a certain Gedanken experiment. For $t > t'$

$$G(\lambda, t > t') = -i\langle|C_\lambda(t)C_\lambda^\dagger(t')|\rangle \tag{2.45}$$

Here one takes the real ground state, and at a time t' one creates an excitation λ. At a later time t one destroys the same excitation. Now if λ were an eigenstate of H, with $HC_\lambda^\dagger|\rangle = \varepsilon_\lambda C_\lambda^\dagger|\rangle$ and $H|\rangle = \varepsilon_0|\rangle$, then this state would propagate with a simple exponential time dependence:

$$G(\lambda, t > t') = -i\ \exp[-i(t - t')(\varepsilon_\lambda - \varepsilon_0)] \tag{2.46}$$

Because λ is not usually an eigenstate of H, the particle in the state λ gets scattered, shifted in energy, etc., during the time interval $t - t'$. When one measures at a later time t, to see how much amplitude is left in the state λ, the measurement provides information about the system.

For the other time arrangement $t' > t$

$$G(\lambda, t' > t) = +i\langle|C_\lambda^\dagger(t')C_\lambda(t)|\rangle \tag{2.47}$$

where the sign in front is changed whenever the position of two fermion operators is interchanged. Now an electron is destroyed from the ground state at time t and recreated at the later time t', which is possible only if there are electrons in the ground state at zero temperature. One case where this is true is in the Fermi sea of a metal. The initial destruction of an electron at time t must, roughly speaking, remove an electron from the filled Fermi sea. This destruction creates a vacancy, often called a *hole*, and the hole can interact and scatter in the interval $t - t'$. Then $C_\lambda^\dagger$ acting at t' destroys the hole state λ, and this measurement provides information about the hole excitation.

The next step is to take the Green's function, defined in the Heisenberg representation, and convert it to the interaction representation. Call $|\rangle_0 \equiv \phi_0$ the ground state of H_0, so that

$$|\rangle = S(0, -\infty)|\rangle_0 \tag{2.48}$$

Next change the operators to this representation:

$$C_\lambda(t) = e^{iHt}e^{-iH_0t}\hat{C}_\lambda(t)e^{iH_0t}e^{-iHt} = U^\dagger(t)\hat{C}_\lambda(t)U(t)$$

$$= S(0,t)\hat{C}_\lambda(t)S(t,0)$$

$$G(\lambda, t-t') = -i\Theta(t-t')_0\langle|S(-\infty,0)S(0,t)\hat{C}_\lambda(t)S(t,0)S(0,t')$$

$$\times \hat{C}^\dagger_\lambda(t')S(t',0)S(0,-\infty)|\rangle_0$$

$$+ i\Theta(t'-t)_0\langle|S(-\infty,0)S(0,t')\hat{C}^\dagger_\lambda(t')S(t',0)S(0,t)$$

$$\times \hat{C}_\lambda(t)S(t,0)S(0,-\infty)|\rangle_0$$

This expression may be regrouped by using the properties of the S matrix developed in the previous section. The left-hand bracket is replaced by

$$_0\langle|S(-\infty,0) = e^{-iL}{}_0\langle|S(\infty,-\infty)S(-\infty,0) = \frac{_0\langle|S(\infty,0)}{_0\langle|S(\infty,-\infty)|\rangle_0}$$

so the Green's function becomes

$$G(\lambda, t-t') = -\frac{i}{_0\langle|S(\infty,-\infty)|\rangle_0}[\Theta(t-t')_0\langle|S(\infty,t)\hat{C}_\lambda(t)$$

$$\times S(t,t')\hat{C}^\dagger_\lambda(t')S(t',-\infty)|\rangle_0$$

$$- \Theta(t'-t)_0\langle|S(\infty,t')\hat{C}^\dagger_\lambda S(t',t)\hat{C}^\dagger_\lambda(t)S(t,-\infty)|\rangle_0]$$

The first term can be simplified by writing

$$\Theta(t-t')_0\langle|S(\infty,t)\hat{C}_\lambda(t)S(t,t')\hat{C}_\lambda(t')S(t',-\infty)|\rangle_0$$

$$= \Theta(t-t')_0\langle|T\hat{C}_\lambda(t)\hat{C}^\dagger_\lambda(t')S(\infty,-\infty)|\rangle_0 \tag{2.49}$$

The operator $S(\infty,-\infty)$ contains operators which act in the three time intervals: (∞, t), (t, t'), and $(t', -\infty)$. The T operator automatically sorts these segments so they act in their proper sequences, which are, respectively, to the left of $\hat{C}_\lambda(t)$, between $\hat{C}_\lambda(t)$ and $\hat{C}^\dagger_\lambda(t')$, and to the right of $\hat{C}^\dagger_\lambda(t')$. The total Green's function is expressed as

$$G(\lambda, t-t') = -i\frac{_0\langle|T\hat{C}_\lambda(t)\hat{C}^\dagger_\lambda(t')S(\infty,-\infty)|\rangle_0}{_0\langle|TS(\infty,-\infty)|\rangle_0} \tag{2.50}$$

which is the desired result. It does not matter where $S(\infty,-\infty)$ is placed in the numerator, since the time ordering operator puts the pieces in the right place.

A Green's function can also be defined for the special case where the interactions $V = 0$ and hence the S matrix is unity. This Green's function plays a special role in the formalism, and it is designated by $G^{(0)}$:

$$G^{(0)}(\lambda, t-t') = -i_0\langle|T\hat{C}_\lambda(t)\hat{C}^\dagger_\lambda(t')|\rangle_0 \tag{2.51}$$

$G^{(0)}$ is often called the *unperturbed Green's function*, or sometimes the name *free propagator* is used.

There are two quite different types of electronic systems in which we want to employ the Green's function analysis. These two have quite different ground states, $|\rangle_0$, and also quite different real ground states, $|\rangle$. These two systems are the following.

1. *An empty band.* Here the properties are studied of an electron in an energy band in which it is the only electron. An example is an electron in the conduction band of a semiconductor or an insulator. In this case the ground state is the particle vacuum, which is denoted as $|0\rangle$. This state has the property that

$$C_{\mathbf{p}}|0\rangle = 0 \tag{2.52}$$

$$a_{\mathbf{q}}|0\rangle = 0 \tag{2.53}$$

where the two destruction operators represent electrons and phonons. Therefore both H_0 and V give zero when operating upon the vacuum. It follows that the S matrix also gives unity when operating upon the vacuum:

$$S(t, -\infty)|0\rangle = |0\rangle \tag{2.54}$$

Both of the ground states, $|\rangle_0$ and $|\rangle$, are the vacuum $|0\rangle$. The Green's function can exist only for the time ordering

$$G(\lambda, t - t') = -i\Theta(t - t')\langle 0|\hat{C}_\lambda(t)\hat{C}^\dagger_\lambda(t')|0\rangle \tag{2.55}$$

The unperturbed Green's function $G^{(0)}$ is particularly easy to evaluate:

$$G^{(0)}(\lambda, t - t') = -i\Theta(t - t')e^{-i\varepsilon_\lambda(t-t')}\langle 0|C_\lambda C^\dagger_\lambda|0\rangle \tag{2.56}$$

$$= -i\Theta(t - t')e^{-i\varepsilon_\lambda(t-t')} \tag{2.57}$$

The Fourier transform of $G(\lambda, t)$ is defined as

$$G(\lambda, E) = \int_{-\infty}^{\infty} dt e^{iEt} G(\lambda, t) \tag{2.58}$$

To make the integrals converge, add the infinitesimal quantity $i\delta$ to the exponents:

$$G^{(0)}(\lambda, E) = -i\int_0^{\infty} dt e^{i(E-\varepsilon_\lambda+i\delta)t}$$

$$G^{(0)}(\lambda, E) = \frac{1}{E - \varepsilon_\lambda + i\delta} \tag{2.59}$$

2. *A degenerate electron gas.* Our second case is where the electrons are in a Fermi sea at zero temperature. The standard example is a simple metal. The system has a chemical potential μ, and all electron states with $E < \mu$ are occupied. If the unperturbed electrons (eigenstates of H_0) are characterized by an energy $\varepsilon_{\mathbf{k}}$, the ground state $|\rangle_0$ has all states $\varepsilon_{\mathbf{k}} < \mu$ filled and states $\varepsilon_{\mathbf{k}} > \mu$ empty. It is convenient and conventional to measure the electron's energy relative to the chemical potential, to define $\xi_{\mathbf{k}} = \varepsilon_{\mathbf{k}} - \mu$. For a spherical Fermi surface with Fermi wave vector p_F,

$${}_0\langle|C^\dagger_{\mathbf{k}}C_{\mathbf{k}}|\rangle_0 = \Theta(p_F - k) \tag{2.60}$$

$${}_0\langle|C_{\mathbf{k}}C^\dagger_{\mathbf{k}}|\rangle_0 = \Theta(k - p_F) \tag{2.61}$$

A more general way to write this result is

$$ {}_0\langle|C_{\mathbf{k}}^{\dagger}C_{\mathbf{k}}|\rangle_0 = \Theta(-\xi_{\mathbf{k}}) = \lim_{\beta\to\infty}\frac{1}{e^{\beta\xi_{\mathbf{k}}}+1} \equiv n_F(\xi_{\mathbf{k}}) $$

The unperturbed Green's function is now

$$ \begin{aligned} G^{(0)}(\lambda, t-t') &= -i\Theta(t-t')_0\langle|C_\lambda(t)C_\lambda^{\dagger}(t')|\rangle_0 + i\Theta(t'-t)_0\langle|C_\lambda^{\dagger}(t')C_\lambda(t)|\rangle_0 \\ &= -i[\Theta(t-t')\theta(\xi_{\mathbf{k}}) - \Theta(t'-t)\Theta(-\xi_{\mathbf{k}})]e^{-i\xi_{\mathbf{k}}(t-t')} \end{aligned} \tag{2.62} $$

The Fourier transform is

$$ \begin{aligned} G^{(0)}(\mathbf{k},\ E) &= -i\left[\theta(\xi_{\mathbf{k}})\int_0^{\infty} dte^{it(E-\xi_{\mathbf{k}}+i\delta)} - \theta(-\xi_{\mathbf{k}})\int_{-\infty}^{0} dte^{it(E-\xi_{\mathbf{k}}-i\delta)}\right] \\ &= \frac{\theta(\xi_{\mathbf{k}})}{E-\xi_{\mathbf{k}}+i\delta} + \frac{\theta(-\xi_{\mathbf{k}})}{E-\xi_{\mathbf{k}}-i\delta} \end{aligned} \tag{2.63} $$

The energy $\xi_{\mathbf{k}} = \varepsilon_{\mathbf{k}} - \mu$ is measured with respect to the Fermi surface $\xi_{\mathbf{k}_F} = 0$. Another way to write $G^{(0)}$ is

$$ G^{(0)}(\mathbf{k}, E) = \frac{1}{E-\xi_{\mathbf{k}}+i\delta_{\mathbf{k}}} \tag{2.64} $$

$$ \delta_{\mathbf{k}} = \delta\ \mathrm{sgn}(\xi_{\mathbf{k}}) \tag{2.65} $$

where $\delta_{\mathbf{k}}$ is a small infinitesimal part which changes sign at the chemical potential $\xi_{\mathbf{k}} = 0$.

3. *Phonons*. The Green's function for phonons is defined as

$$ D(\mathbf{q}, \lambda, t-t') = -i\langle|TA_{\mathbf{q}\lambda}(t)A_{-\mathbf{q}\lambda}(t')|\rangle \tag{2.66} $$

$$ A_{\mathbf{q}\lambda} = a_{\mathbf{q}\lambda} + a_{-\mathbf{q}\lambda}^{\dagger} \tag{2.67} $$

The subscripts λ refer to the polarization of the phonons. Most applications have one kind of phonon in Hamiltonians which do not mix polarizations, and then these subscripts are omitted entirely. In the interaction representation

$$ D(\mathbf{q}, t-t') = -i\frac{{}_0\langle|T\hat{A}_{\mathbf{q}}(t)\hat{A}_{-\mathbf{q}}(t')S(\infty,-\infty)|\rangle_0}{{}_0\langle|S(\infty,-\infty)|\rangle_0} \tag{2.68} $$

At zero temperature there are no phonons. The ground states $|\rangle$ and $|\rangle_0$ are again the particle vacuum $|0\rangle$. Note that in an electron–phonon system the notation $|\rangle_0$ means the combination of ground states for electrons, phonons, etc. Although the phonon system has the vacuum as its ground state, either of the two-electron ground states can be used.

The unperturbed phonon Green's function is defined as

$$ \begin{aligned} D^{(0)}(\mathbf{q}, t-t') &= -i_0\langle|T\hat{A}_{\mathbf{q}}(t)\hat{A}_{-\mathbf{q}}(t')|\rangle_0 \\ &= -i_0\langle|T(a_{\mathbf{q}}e^{-i\omega_{\mathbf{q}}t} + a_{-\mathbf{q}}^{\dagger}e^{i\omega_{\mathbf{q}}t})(a_{-\mathbf{q}}e^{-i\omega_{\mathbf{q}}t'} + a_{\mathbf{q}}^{\dagger}e^{i\omega_{\mathbf{q}}t'})|\rangle_0 \end{aligned} \tag{2.69} $$

At zero temperature

$$_0\langle|a_{\mathbf{q}}a_{\mathbf{q}}^\dagger|\rangle_0 = 1$$

$$_0\langle|a_{\mathbf{q}}^\dagger a_{\mathbf{q}}|\rangle_0 = 0$$

$$D^{(0)}(\mathbf{q}, t-t') = -i[\Theta(t-t')e^{-i\omega_{\mathbf{q}}(t-t')} + \Theta(t'-t)e^{i\omega_{\mathbf{q}}(t-t')}] \tag{2.70}$$

The Fourier transform has two terms which correspond to the two time orderings:

$$D^{(0)}(\mathbf{q}, \omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} D^{(0)}(\mathbf{q}, t) \tag{2.71}$$

$$D^{(0)}(\mathbf{q}, \omega) = \frac{1}{\omega - \omega_{\mathbf{q}} + i\delta} - \frac{1}{\omega + \omega_{\mathbf{q}} - i\delta} \tag{2.72}$$

$$= \frac{2\omega_{\mathbf{q}}}{\omega^2 - \omega_{\mathbf{q}}^2 + i\delta} \tag{2.73}$$

Sometimes it is useful to have the phonon Green's function at nonzero temperature. In this case, the thermal average is taken of the phonon occupation numbers,

$$_0\langle|a_{\mathbf{q}}a_{\mathbf{q}}^\dagger\rangle_0 = N_{\mathbf{q}} + 1$$

$$_0\langle|a_{\mathbf{q}}^\dagger a_{\mathbf{q}}|\rangle_0 = N_{\mathbf{q}} = \frac{1}{e^{\beta\omega_{\mathbf{q}}} - 1}$$

and the Green's function of time is

$$D^{(0)}(\mathbf{q}, t-t') = -i[(N_{\mathbf{q}} + 1)e^{-i\omega_{\mathbf{q}}|t-t'|} + N_{\mathbf{q}}e^{i\omega_{\mathbf{q}}|t-t'|}] \tag{2.74}$$

In some many-body systems the interactions cause changes in the phonon energies. For these cases, the thermal average at nonzero temperature should be over the frequencies which result from the interactions. The use of the unperturbed frequencies is valid only if they are not altered, which happens when the perturbation is either localized or confined to a small number of particles.

2.4. WICK'S THEOREM

The Green's function is evaluated by expanding the S matrix $S(\infty, -\infty)$ in (2.33) in a series such as (2.26):

$$G(\mathbf{p}, t-t') = \sum_{n=0}^{\infty} \frac{(-i)^{n+1}}{n!} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n \tag{2.75}$$

$$\times \frac{_0\langle|T\hat{C}_{\mathbf{p}}(t)\hat{V}(t_1)\hat{V}(t_2)\cdots\hat{V}(t_n)\hat{C}_{\mathbf{p}}^\dagger(t')|\rangle_0}{_0\langle|S(\infty, -\infty)|\rangle_0} \tag{2.76}$$

Let us, for the moment, ignore the phase factor $_0\langle|S(\infty, -\infty)|\rangle_0$. It will be taken care of in Sec. 2.6. The immediate aim is to learn how to evaluate time-ordered brackets such as

$$_0\langle|T\hat{C}_{\mathbf{p}}(t)\hat{V}(t_1)\hat{V}(t_2)\hat{V}(t_3)\hat{C}_{\mathbf{p}}^\dagger(t')|\rangle_0 \tag{2.77}$$

Suppose that $\hat{V}(t_1)$ is the electron–electron interaction Eq. (1.160):

$$\hat{V}(t_1) = \frac{1}{2}\sum_{\mathbf{k}'\mathbf{k}\mathbf{q}}\sum_{ss'}\frac{4\pi e^2}{q^2}C^\dagger_{\mathbf{k}+\mathbf{q},s}C^\dagger_{\mathbf{k}'-\mathbf{q},s'}C_{\mathbf{k}',s'}C_{\mathbf{k},s}e^{it_1(\xi_{\mathbf{k}+\mathbf{q}}+\xi_{\mathbf{k}'-\mathbf{q}}-\xi_{\mathbf{k}}-\xi_{\mathbf{k}'})}$$

In this case the time-ordered bracket (2.77) contains seven creation operators and seven destruction operators. It is a very arduous task to evaluate this bracket: there are many possible time orderings and many possible pairings between creation and destruction operators. What is meant by *pairings between operators*? First note that these brackets always contain the same number of creation and destruction operators. One is always trying to evaluate the product of n creation operators and n destruction operators between the ground state $|\rangle_0$

$$_0\langle|T\hat{C}_1(t)\hat{C}^\dagger_{1'}(t_1)\cdots\hat{C}_n(t_n)\hat{C}^\dagger_{n'}(t')|\rangle_0 \tag{2.78}$$

The effect of a creation operator $\hat{C}^\dagger_{n'}(t')$ is to put an electron into the state n'. The system must be back in the ground state before the final operator of $_0\langle|$, so that one of the destruction operators $\hat{C}_m(t_m)$ must destroy the state n' and $m = n'$ for some m. For example,

$$_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\beta(t')|\rangle_0 \tag{2.79}$$

equals zero unless $\alpha = \beta$, while

$$_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\beta(t_1)\hat{C}_\gamma(t_2)\hat{C}^\dagger_\delta(t')|\rangle_0 \tag{2.80}$$

equals zero unless $\alpha = \beta, \gamma = \delta$, or unless $\alpha = \delta$, $\beta = \gamma$. There are many possible time orderings and pairings in a bracket like (2.77). However, only a limited number of these combinations are physically interesting. Our aim is to sort these in a simple way to identify the important terms. This sorting is achieved with the help of some theorems which simplify the procedures. The first of these is *Wick's theorem.*

This theorem is really just an observation that the time ordering can be taken care of in a simple way. Wick's theorem states that, in making all the possible pairings between creation and destruction operators, each pairing should be time-ordered. The time ordering of each pair gives the proper time ordering to the entire result.

For example,

$$\begin{aligned}
&_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\beta(t_1)\hat{C}_\gamma(t_2)\hat{C}^\dagger_\delta(t')|\rangle_0\\
&= {}_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\beta(t_1)|\rangle_{00}\langle|T\hat{C}_\gamma(t_2)\hat{C}^\dagger_\delta(t')|\rangle_0\\
&\quad - {}_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\delta(t')|\rangle_{00}\langle|T\hat{C}_\gamma(t_2)\hat{C}^\dagger_\beta(t_1)|\rangle_0\\
&= \delta_{\alpha\beta}\delta_{\gamma\delta}\,{}_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\alpha(t_1)|\rangle_{00}\langle|T\hat{C}_\gamma(t_2)\hat{C}^\dagger_\gamma(t')|\rangle_0\\
&\quad - \delta_{\alpha\delta}\delta_{\beta\gamma}\,{}_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\alpha(t')|\rangle_{00}\langle|T\hat{C}_\gamma(t_2)\hat{C}^\dagger_\gamma(t_1)|\rangle_0
\end{aligned}$$

Note that there is a time-ordering operator T in each of the two pairing brackets. For $n = 3$ creation and destruction operators in the original bracket, there are six possible pairings; for n operators of each kind there are $n!$ possible pairings. Also note that within a pairing bracket, the labels α, β, etc., must be the same. These labels denote eigenstates, so the creation and destruction operators must refer to the same state.

A few simple rules should be kept in mind when making these pairings. The first is that a sign change occurs each time the positions of two neighboring Fermi operators are interchanged. One keeps count of the number of interchanges needed to achieve the desired pairing. An odd number of interchanges is the origin of the minus sign in the second term of the example above.

The second rule concerns the time ordering of combinations of operators representing different excitations. For example, consider the following mixture of phonon and electron operators:

$$_0\langle|T\hat{C}_{\mathbf{p}}(t)\hat{C}^{\dagger}_{\mathbf{p}_1}(t_1)\hat{A}_{\mathbf{q}_1}(t_1)\hat{C}_{\mathbf{p}_2}(t_2)\hat{C}^{\dagger}_{\mathbf{p}_3}(t_3)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_0 \tag{2.81}$$

Because electron operators commute with phonon operators, it is not important how they are ordered with respect to each other. The bracket can be immediately factored into separate parts for electrons and phonons:

$$_0\langle|T\hat{C}_{\mathbf{p}}(t)\hat{C}^{\dagger}_{\mathbf{p}_1}(t_1)\hat{C}_{\mathbf{p}_2}(t_2)\hat{C}^{\dagger}_{\mathbf{p}_3}(t_3)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_0 \tag{2.82}$$

This separation is always possible with different kinds of operators, i.e., whenever operators commute. Wick's theorem also applies to brackets of phonon operators; for example,

$$\begin{aligned}
&_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)\hat{A}_{\mathbf{q}_3}(t_3)\hat{A}_{\mathbf{q}_4}(t_4)|\rangle_0 \\
&= {}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_3}(t_3)\hat{A}_{\mathbf{q}_4}(t_4)|\rangle_0 \\
&\quad + {}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_3}(t_3)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_2}(t_2)\hat{A}_{\mathbf{q}_4}(t_4)|\rangle_0 \\
&\quad + {}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_4}(t_4)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_2}(t_2)\hat{A}_{\mathbf{q}_3}(t_3)|\rangle_0
\end{aligned} \tag{2.83}$$

In this case, for boson operators, there is no change of sign when the order of the operators is changed. Each of the above brackets vanishes unless the two wave vectors are equal and opposite:

$$\begin{aligned}
&= \delta_{\mathbf{q}_1+\mathbf{q}_2}\delta_{\mathbf{q}_3+\mathbf{q}_4}\,{}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{-\mathbf{q}_1}(t_2)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_3}(t_3)\hat{A}_{-\mathbf{q}_3}(t_4)|\rangle_0 \\
&\quad + \delta_{\mathbf{q}_1+\mathbf{q}_3}\delta_{\mathbf{q}_2+\mathbf{q}_4}\,{}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{-\mathbf{q}_1}(t_3)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_2}(t_2)\hat{A}_{-\mathbf{q}_2}(t_4)|\rangle_0 \\
&\quad + \delta_{\mathbf{q}_1+\mathbf{q}_4}\delta_{\mathbf{q}_2+\mathbf{q}_3}\,{}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{-\mathbf{q}_1}(t_4)|\rangle_{00}\langle|T\hat{A}_{\mathbf{q}_2}(t_2)\hat{A}_{-\mathbf{q}_2}(t_3)|\rangle_0
\end{aligned} \tag{2.84}$$

The third rule is a method of treating the "time ordering" of two operators which occur at the same time, such as

$$_0\langle|T\hat{C}^{\dagger}_{\mathbf{k}_1}(t_1)\hat{C}_{\mathbf{k}_2}(t_1)|\rangle_0 \tag{2.85}$$

In these cases the destruction operator always goes to the right:

$$= \delta_{\mathbf{k}_1=\mathbf{k}_2}\,{}_0\langle|\hat{C}^{\dagger}_{\mathbf{k}_1}(t_1)\hat{C}_{\mathbf{k}_1}(t_1)|\rangle_0 = \delta_{\mathbf{k}_1=\mathbf{k}_2}n_F(\xi_{\mathbf{k}_1}) \tag{2.86}$$

and the term is just the number operator which is independent of time. This convention is dependent on the convention used to write down the Hamiltonian. In constructing H the destruction operators are put to the right of the creation operators in all terms in the Hamiltonian.

When two electron operators have *different time arguments* in a pairing, it is conventional to put the creation operator to the right:

$$ {}_0\langle|T\hat{C}^\dagger_{\mathbf{k}_1}(t_1)\hat{C}_{\mathbf{k}_2}(t_2)|\rangle_0 = -\delta_{\mathbf{k}_1=\mathbf{k}_2}\,{}_0\langle|T\hat{C}_{\mathbf{k}_1}(t_2)\hat{C}^\dagger_{\mathbf{k}_1}(t_1)|\rangle_0 \tag{2.87} $$

This term can be immediately identified as the unperturbed Green's function $iG^{(0)}(\mathbf{k}_1, t_2 - t_1)$. All of the pairing brackets for electron operators are either Green's functions or else number operators. The previous examples can also be written in terms of Green's functions:

$$ \begin{aligned} &(-i)^2\,{}_0\langle|T\hat{C}_\alpha(t)\hat{C}^\dagger_\beta(t_1)\hat{C}_\gamma(t_2)\hat{C}^\dagger_\delta(t')|\rangle_0 \\ &= \delta_{\alpha\beta}\delta_{\gamma\delta}G^{(0)}(\alpha, t-t_1)G^{(0)}(\gamma, t_2 - t') \\ &\quad - \delta_{\alpha\delta}\delta_{\beta\gamma}G^{(0)}(\alpha, t-t')G^{(0)}(\gamma, t_2 - t_1) \\ &(-i)^3\,{}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)\hat{A}_{\mathbf{q}_3}(t_3)\hat{A}_{\mathbf{q}_4}(t_4)|\rangle_0 \\ &= \delta_{\mathbf{q}_1+\mathbf{q}_2}\delta_{\mathbf{q}_3+\mathbf{q}_4}D^{(0)}(\mathbf{q}_1, t_1 - t_2)D^{(0)}(\mathbf{q}_3, t_3 - t_4) \\ &\quad + \delta_{\mathbf{q}_1+\mathbf{q}_3}\delta_{\mathbf{q}_2+\mathbf{q}_4}D^{(0)}(\mathbf{q}_1, t_1 - t_3)D^{(0)}(\mathbf{q}_2, t_2 - t_4) \\ &\quad + \delta_{\mathbf{q}_1+\mathbf{q}_4}\delta_{\mathbf{q}_2+\mathbf{q}_3}D^{(0)}(\mathbf{q}_1, t_1 - t_4)D^{(0)}(\mathbf{q}_2, t_2 - t_3) \end{aligned} $$

In summary, Wick's theorem states that a time-ordered bracket may be evaluated by expanding it into all possible pairings. Each of these pairings will be a time-ordered Green's function or a number operator n_F or n_B. This expansion gets the correct time-ordering for the entire brackets. Wick's theorem is valid only when the Hamiltonian H_0 is bilinear in creation and destruction operators. The theorem fails if H_0 contains pair-wise interactions between particles.

A comprehensive example is done for the $n = 2$ term in the S matrix expansion in (2.76). The $n = 0$ term is always $G^{(0)}$:

$$ \begin{aligned} G(\mathbf{p}, t-t') = G^{(0)}(\mathbf{p}, t-t') &+ (-i)^2\int_{-\infty}^{\infty} dt_1\ {}_0\langle|T\hat{C}_{\mathbf{p}\sigma}(t)\hat{V}(t_1)\hat{C}^\dagger_{\mathbf{p}\sigma}(t')|\rangle_0 \\ &+ \frac{(-i)^3}{2!}\int_{-\infty}^{\infty} dt_1\int_{-\infty}^{\infty} dt_2\ {}_0\langle|T\hat{C}_{\mathbf{p}\sigma}(t)\hat{V}(t_1)\hat{V}(t_2)\hat{C}^\dagger_{\mathbf{p}\sigma}(t')|\rangle_0 \end{aligned} $$

The interaction will be taken as the electron–phonon interaction:

$$ V = \sum_{\mathbf{q}\mathbf{k}s} M_{\mathbf{q}}A_{\mathbf{q}}C^\dagger_{\mathbf{k}+\mathbf{q},s}C_{\mathbf{k}s} \tag{2.88} $$

First note that the $n = 1$ term must vanish because it contains the factor

$$ {}_0\langle|T\hat{A}_{\mathbf{q}}|\rangle_0 \tag{2.89} $$

which is zero since the factors $\langle 0|a_{\mathbf{q}}|0\rangle$ and $\langle 0|a^\dagger_{\mathbf{q}}|0\rangle$ are zero. Similarly, all the terms where n is odd vanish because their time-ordered bracket for phonons contains an odd number of $\hat{A}_{\mathbf{q}}$

factors. Only the terms even in n contribute to the S-matrix expansion for the electron–phonon interaction:

$$
\begin{aligned}
G(\mathbf{p}, t-t') = G^{(0)}(\mathbf{p}, t-t') &+ \frac{(-i)^3}{2!}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \\
&\times \sum_{\mathbf{q}_1\mathbf{q}_2} M_{\mathbf{q}_1} M_{\mathbf{q}_2}\, {}_0\langle | T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_0 \\
&\times \sum_{\mathbf{k}_1\mathbf{k}_2 ss'} {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_{1,s}}(t_1)\hat{C}_{\mathbf{k}_{1,s}}(t_1) \\
&\times \hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_{2,s'}}(t_2)\hat{C}_{\mathbf{k}_{2s'}}(t_2)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0
\end{aligned}
$$

The phonon bracket gives a single-phonon Green's function:

$$
{}_0\langle | T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_0 = i\delta_{\mathbf{q}_1+\mathbf{q}_2} D^{(0)}(\mathbf{q}_1, t_1 - t_2) \tag{2.90}
$$

The electron bracket, unfortunately, has six possible combinations of pairings. The six terms are evaluated using the fact that $\mathbf{q}_2 = -\mathbf{q}_1$. Wick's theorem gives the result

$$
\begin{aligned}
&{}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)\hat{C}_{\mathbf{k}_1,s}(t_1)\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)\hat{C}_{\mathbf{k}_2 s'}(t_2)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0 \\
&= {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)|\rangle_0 {}_0\langle | T\hat{C}_{\mathbf{k}_1,s}(t_1)\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)|\rangle_0 \\
&\quad \times {}_0\langle | T\hat{C}_{\mathbf{k}_2 s'}(t_2)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0 \\
&\quad + {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)|\rangle_0 {}_0\langle | T\hat{C}_{\mathbf{k}_1,s}(t_1)\hat{C}^{\dagger}_{\mathbf{p},\sigma}(t')|\rangle_0 \\
&\quad \times {}_0\langle | T\hat{C}_{\mathbf{k}_2 s'}(t_2)\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)|\rangle_0 \\
&\quad + {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q},s}(t_1)|\rangle_0 {}_0\langle | T\hat{C}_{\mathbf{k}_1,s}(t_1)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0 \\
&\quad \times {}_0\langle | T\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)\hat{C}_{\mathbf{k}_2 s'}(t_2)|\rangle_0 \\
&\quad + {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)|\rangle_0 {}_0\langle | T\hat{C}_{\mathbf{k}_2,s'}(t_2)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0 \\
&\quad \times {}_0\langle | T\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)\hat{C}_{\mathbf{k}_1 s}(t_1)|\rangle_0 \\
&\quad + {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0 {}_0\langle | \hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)\hat{C}_{\mathbf{k}_1,s}(t_1)|\rangle_0 \\
&\quad \times {}_0\langle | T\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)\hat{C}_{\mathbf{k}_2,s'}(t_2)|\rangle_0 \\
&\quad - {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')|\rangle_0 {}_0\langle | T\hat{C}_{\mathbf{k}_1,s}(t_1)\hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)|\rangle_0 \\
&\quad \times {}_0\langle | T\hat{C}_{\mathbf{k}_2 s'}(t_2)\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)|\rangle_0
\end{aligned}
$$

Each bracket is either $G^{(0)}$ functions with different time arguments, or else a number operator n_F at equal times. The six terms, taken in the same order, give

$$
\begin{aligned}
& i^3 \delta_{\mathbf{p}=\mathbf{k}_2=\mathbf{k}_1+\mathbf{q}_1} \delta_{s=s'=\sigma} G^{(0)}(\mathbf{p}, t-t_1) G^{(0)}(\mathbf{p}-\mathbf{q}_1, t_1-t_2) G^{(0)}(\mathbf{p}, t_2-t') \\
& \quad + i^2 \delta_{\mathbf{p}=\mathbf{k}_1=\mathbf{k}_2-\mathbf{q}_1} \delta_{s=s'=\sigma} G^{(0)}(\mathbf{p}, t-t_2) G^{(0)}(\mathbf{p}, +\mathbf{q}_1, t_2-t_1) G^{(0)}(\mathbf{p}, t_1-t') \\
& \quad + i^2 \delta_{\mathbf{q}_1=0} \delta_{\mathbf{p}=\mathbf{k}_1} \delta_{s=\sigma} n_F(\xi_{\mathbf{k}_2}) G^{(0)}(\mathbf{p}, t-t_1) G^{(0)}(\mathbf{p}, t_1-t') \\
& \quad + i^2 \delta_{\mathbf{q}_1=0} \delta_{\mathbf{p}=\mathbf{k}_2} \delta_{s'=\sigma} n_F(\xi_{\mathbf{k}_1}) G^{(0)}(\mathbf{p}, t-t_2) G^{(0)}(\mathbf{p}, t_2-t') \\
& \quad + i \delta_{\mathbf{q}_1=0} \delta_{\mathbf{q}_2=0} n_F(\xi_{\mathbf{k}_1}) n_F(\xi_{\mathbf{k}_2}) G^{(0)}(\mathbf{p}, t-t') \\
& \quad - i^3 \delta_{\mathbf{k}_1=\mathbf{k}_2-\mathbf{q}_1} \delta_{s'=s} G^{(0)}(\mathbf{p}, t-t') G^{(0)}(\mathbf{k}_1, t_1-t_2) G^{(0)}(\mathbf{k}_1+\mathbf{q}_1, t_2-t_1)
\end{aligned}
\tag{2.91}
$$

2.5. FEYNMAN DIAGRAMS

Feynman introduced the idea of representing the kind of terms in (2.91) by drawings. These drawings, called diagrams, are extremely useful for providing an insight into the physical process which these terms represent. These diagrams can be drawn both for the Green's function depending on time $G(\mathbf{p}, t)$ as well as for functions which are Fourier-transformed and depend on energy $G(\mathbf{p}, E)$.

The diagrams in time are drawn by representing the electron Green's function $G^{(0)}(\mathbf{p}, t-t')$ by a solid line which goes from t' to t, as shown in Fig. 2.1. An arrow is often included to represent the direction. The arrow is mostly for convenience, and it does not imply or require that $t > t'$. The phonon Green's function is represented by a dashed line. It does not have a directional arrow because

$$D^{(0)}(\mathbf{q}, t-t') = D^{(0)}(-\mathbf{q}, t'-t) \tag{2.92}$$

and the sign of $\mathbf{q}$ is irrelevant. Phonons can be viewed as going in either direction in time. Next consider the diagram for the factor

$${}_0\langle | C^\dagger_{\mathbf{p}s}(t) C_{\mathbf{p}s}(t) | \rangle_0 = n_F(\xi_{\mathbf{p}}) \tag{2.93}$$

The occupation number n_F is drawn as a solid line which loops and represents an electron line which starts and ends at the same point in time. By using these rules, the diagrams are constructed which represent each of the six terms in the S-matrix expansion (2.91). These six

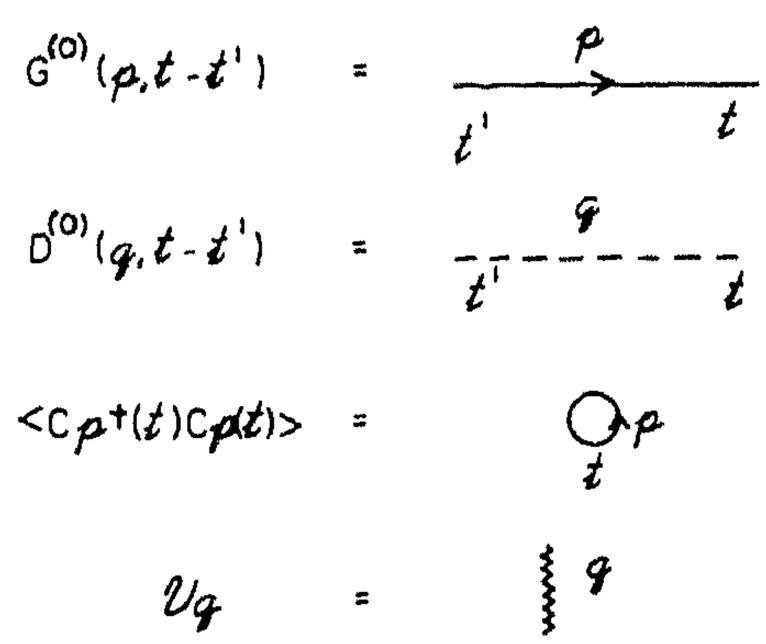


FIGURE 2.1

terms are shown in Fig. 2.2, and the diagrams (a) to (f) are the six terms in (2.91) in the same order. Each term has the phonon line connecting the times t_1 and t_2.

In Fig. 2.2 the terms (c), (d), and (e) are zero. They vanish because they exist only if the phonon wave vector $\mathbf{q}$ is zero, but there are no phonons with $\mathbf{q} = 0$. A phonon with $\mathbf{q} = 0$ is either a translation of the crystal or a permanent strain, and neither of these is meant to be in the Hamiltonian. The sum over $\mathbf{q}$ in (2.88) should exclude the $\mathbf{q} = 0$ term.

The two terms (a) and (b) in Fig. 2.2 are not zero. They are the contributions of primary interest. By using the results of (2.91), the contributions of the terms (a) and (b) are

$$\begin{aligned}\frac{1}{2!}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \sum_{\mathbf{q}} |M_{\mathbf{q}}|^2 D^{(0)}(\mathbf{q}, t_1 - t_2)\\ \times G^{(0)}(\mathbf{p}, t - t_1)G^{(0)}(\mathbf{p} - \mathbf{q}, t_1 - t_2)G^{(0)}(\mathbf{p}, t_2 - t')\\ + G^{(0)}(\mathbf{p}, t - t_2)G^{(0)}(\mathbf{p} + \mathbf{q}, t_2 - t_1)G^{(0)}(\mathbf{p}, t_1 - t')]\end{aligned} \tag{2.94}$$

The two drawings (a) and (b) look alike. They differ only in the labeling of the variables $t_1, t_2, \mathbf{q}$. But these are variables of integration in (2.94) and may be relabeled which shows the two terms are equal. Only one term need be written, and the factor of $1/2!$ in front is eliminated.

The term (f) in Fig. 2.2 is

$$G^{(0)}(\mathbf{p}, t - t')F_1 \tag{2.95}$$

$$\begin{aligned}F_1 = -\frac{i}{2!}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \sum_{\mathbf{kq}} |M_{\mathbf{q}}|^2 D^{(0)}(\mathbf{q}, t_1 - t_2)G^{(0)}(\mathbf{k}, t_1 - t_2)\\ \times G^{(0)}(\mathbf{k} + \mathbf{q}, t_2 - t_1)\end{aligned} \tag{2.96}$$

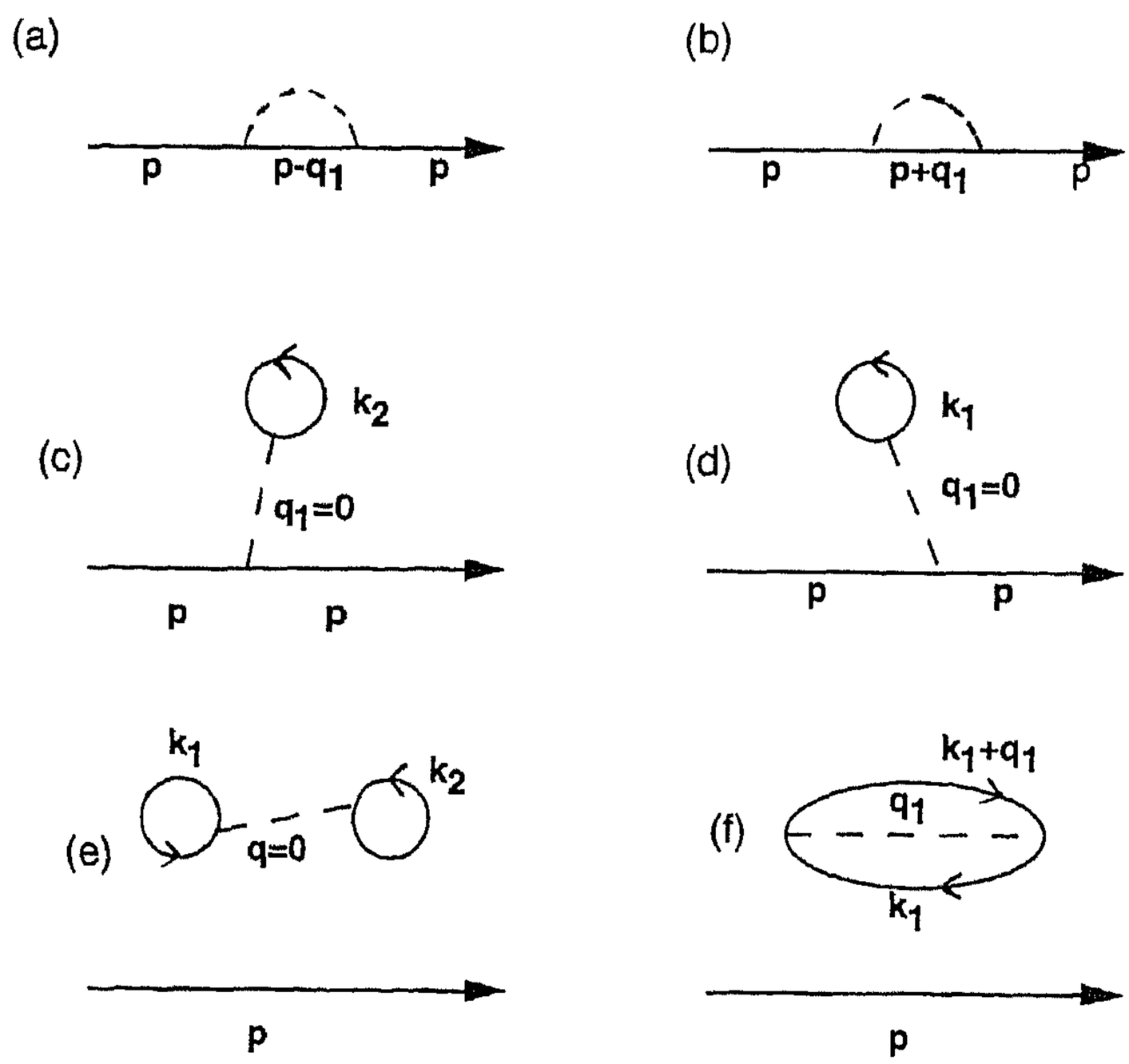


FIGURE 2.2

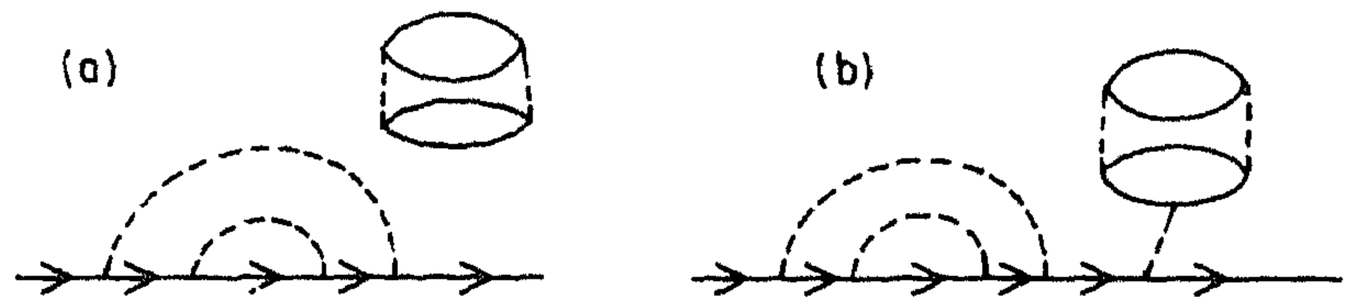


FIGURE 2.3

This drawing has the property that part of it is not topologically connected to the Green's function line $G^{(0)}(\mathbf{p}, t-t')$. Diagrams in which *all* parts are not connected are called *disconnected diagrams*. For example, in Fig. 2.3, part (a) is disconnected, while part (b) is connected. The disconnected parts, as in (2.95), provide just a multiplicative constant such as F_1 which multiplies the contribution from the connected parts.

2.6. VACUUM POLARIZATION GRAPHS

Now consider the factor which has been ignored to this point:

$$
{}_0\langle|S(\infty,-\infty|\rangle_0 = \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\cdots\int_{-\infty}^{\infty}dt_n
$$

$$
\times\, {}_0\langle|T\hat{V}(t_1)\hat{V}(t_2)\cdots\hat{V}(t_n)|\rangle_0
$$

Again consider the electron–phonon interaction and evaluate the term $n=2$. The $n=1$ term vanishes as it did for the Green's function expansion,

$$
{}_0\langle|S|\rangle_0 = 1 + \frac{(-i)^2}{2!}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\ {}_0\langle|T\hat{V}(t_1)\hat{V}(t_2)|\rangle_0 \tag{2.97}
$$

where

$$
\begin{aligned}
{}_0\langle|T\hat{V}(t_1)\hat{V}(t_2)|\rangle_0 = \sum_{\mathbf{q}_1\mathbf{q}_2} M_{\mathbf{q}_1}M_{\mathbf{q}_2}\ {}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_0 \\
\times \sum_{\mathbf{k}_1\mathbf{k}_2 ss'}\ {}_0\langle|T\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)\hat{C}_{\mathbf{k}_1 s}(t_1) \\
\times \hat{C}^{\dagger}_{\mathbf{k}_2+\mathbf{q}_2,s'}(t_2)\hat{C}_{\mathbf{k}_2,s'}(t_2)|\rangle_0
\end{aligned}
$$

By using Wick's theorem,

$$
\begin{aligned}
&{}_0\langle|T\hat{A}_{\mathbf{q}_1}(t_1)\hat{A}_{\mathbf{q}_2}(t_2)|\rangle_0 = i\delta_{\mathbf{q}_1+\mathbf{q}_2}D^{(0)}(\mathbf{q}_1, t_1-t_2) \\
&{}_0\langle|T\hat{C}^{\dagger}_{\mathbf{k}_1+\mathbf{q}_1,s}(t_1)\hat{C}_{\mathbf{k}_1 s}(t_1)\hat{C}^{\dagger}_{\mathbf{k}_2-\mathbf{q}_1,s'}(t_2)\hat{C}_{\mathbf{k}_2,s'}(t_2)|\rangle_0 \\
&\quad = \delta_{\mathbf{q}_1}n_F(\xi_{\mathbf{k}_1})n_F(\xi_{\mathbf{k}_2}) + \delta_{\mathbf{k}_1=\mathbf{k}_2-\mathbf{q}_1}G^{(0)}(\mathbf{k}_1, t_1-t_2)G^{(0)}(\mathbf{k}_1+\mathbf{q}_1, t_2-t_1)
\end{aligned} \tag{2.98}
$$

The Feynman diagrams for the two terms in Eq. (2.98) are shown in Figs. 2.4(a) and 2.4(b). The (a) term is zero because there are no $\mathbf{q}=0$ phonons. The (b) term is nonzero and gives a contribution

$$
{}_0\langle|S(\infty,-\infty|\rangle_0 = 1 + F_1 + \cdots \tag{2.99}
$$

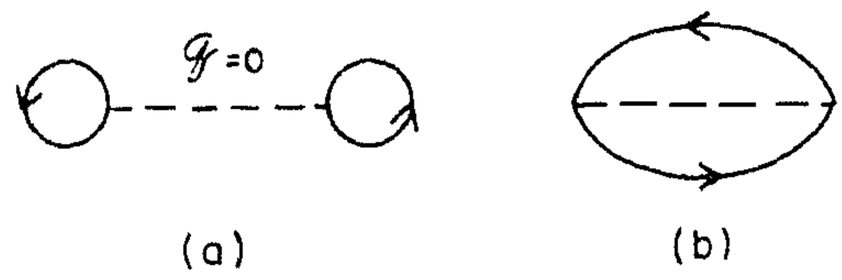


FIGURE 2.4

where F_1 is defined in (2.96). The constant F_1 appears whenever the closed bubble of Fig. 2.4(b) occurs, regardless of whether the term arises in the disconnected diagrams of $G(\mathbf{p}, t-t')$ or in the expansion of ${}_0\langle|S(\infty,-\infty|\rangle_0$.

The terms in the series for ${}_0\langle|S(\infty,-\infty|\rangle_0$ are called *vacuum polarization terms*. Some terms for $n=4$, where there are two phonon lines, are shown in Fig. 2.5. Each of these diagrams or terms represents a constant F_j which one can evaluate by doing the required time and wave vector integrals. The constant ${}_0\langle|S(\infty,-\infty|\rangle_0$ could be evaluated by computing all the F_j and then summing them ($F_0=1$):

$$ {}_0\langle|S(\infty,-\infty|\rangle_0 = \sum_{j=0}^{\infty} F_j \tag{2.100}$$

This procedure is unnecessary because of a cancellation theorem.

The next theorem also simplifies the calculation of the Green's function expansion (2.76). This theorem is that the vacuum polarization diagrams exactly cancel the disconnected diagrams in the expansion for $G(\mathbf{p}, t-t')$. The net result is that in calculating $G(\mathbf{p}, t-t')$ one needs only to evaluate the connected diagrams. The other contributions, from the disconnected diagrams and from ${}_0\langle|S(\infty,-\infty|\rangle_0$, exactly cancel one another.

This theorem will not be proven, but only explained. Call $G_c(\mathbf{p}, t-t')$ the summation of all connected diagrams, and the basic theorem is that

$$ {}_0\langle|T\hat{C}_{\mathbf{p}}(t)\hat{C}^{\dagger}_{\mathbf{p}}(t')S(\infty,-\infty|\rangle_0 = G_c(\mathbf{p}, t-t')\,{}_0\langle|S(\infty,-\infty|\rangle_0$$

The Green's function (2.43) is just the summation of all the connected diagrams:

$$G(\mathbf{p}, t-t') = G_c(\mathbf{p}, t-t') \tag{2.101}$$

The proof of this theorem is just a counting problem. One must convince oneself that each connected diagram has, in higher-order terms in the S-matrix expansion, all disconnected parts which exactly add up to ${}_0\langle|S(\infty,-\infty|\rangle_0$. For example, the self-energy diagram in Fig. 2.2(a) has in higher order the vacuum polarization terms shown in Fig. 2.6. The summation of all these terms, to all orders, is just the factor ${}_0\langle|S(\infty,-\infty|\rangle_0$. The important point is that each disconnected part is just a constant factor F_j. This theorem is very convenient, since it states that one can just ignore the disconnected diagrams. They do not need to be calculated. It is just as well, since when they are evaluated they often turn out to be infinity. In fact, it is easy

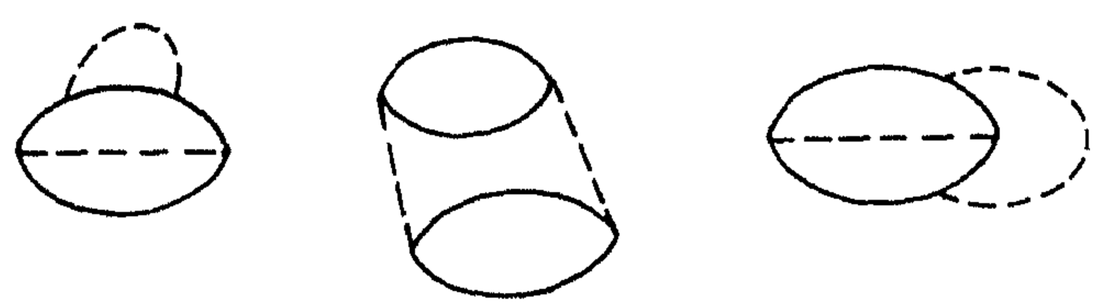


FIGURE 2.5 Vacuum polarization graphs

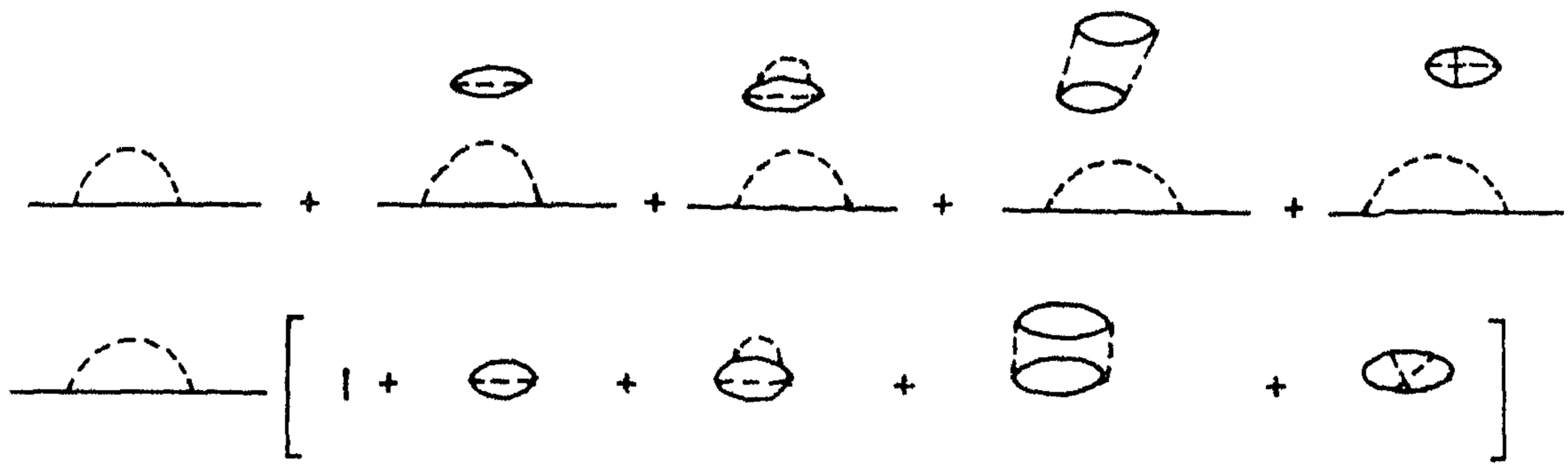


FIGURE 2.6

to show that they are infinity. The disconnected term F_1 is defined in (2.96). The integrand is only a function $f(t_1 - t_2)$ of $t_1 - t_2$. Change integration variables to

$$\tau = t_1 - t_2$$

$$s = \tfrac{1}{2}(t_1 + t_2)$$

and Eq. (2.96) becomes

$$F_1 = \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} d\tau f(\tau) \tag{2.102}$$

The important point is that there is no dependence on s, so

$$\int_{-\infty}^{\infty} ds = \infty \tag{2.103}$$

One can show that each disconnected part has an "extra" time integral and has the same infinity.

It has been shown that the one-particle Green's functions consist of just connected diagrams:

$$G(\mathbf{p}, t - t') = -i \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n$$
$$\times\, {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t_1)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')\hat{V}(t_1)\cdots\hat{V}(t_n)|\rangle_0 \text{ (connected)}$$

The next step is to get rid of the $1/n!$ factor. It is eliminated because there are just $n!$ terms exactly alike in each bracket of the nth term in the expansion.

Considering only different terms gives the result

$$G(\mathbf{p}, t - t') = -i \sum_{n=0}^{\infty} (-i)^n \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t)\hat{C}^{\dagger}_{\mathbf{p}\sigma}(t')$$
$$\times\, \hat{V}(t_1)\cdots\hat{V}(t_n)|\rangle_0 \text{ (different connected)} \tag{2.104}$$

The obvious question is then: how can you tell when terms are different? Usually one can tell by inspection, although sometimes terms must be examined carefully. In Fig. 2.2, terms (a) and (b) are the same and provide the $2 = 2!$ necessary for this $n = 2$ term. Similarly, (c) and (e) are identical. For example, (a) and (b) differ only in the variables t_1, t_2 and $\mathbf{q}_1$. But these are dummy variables of integration and so may be relabeled, respectively, to t_2, t_1 and $-\mathbf{q}_1$. Then the (a) term is obviously the same as (b).

The next terms in the electron–phonon expansion have $n = 4$, which are diagrams with two phonons. Here each different connected diagram is found $4! = 24$ times.

2.7. DYSON'S EQUATION

The Green's function of energy is defined by taking the usual Fourier transform with respect to the time variable:

$$G(\mathbf{p}, E) = \int_{-\infty}^{\infty} dt e^{iE(t-t')} G(\mathbf{p}, t - t') \tag{2.105}$$

This time integral has already been evaluated for the unperturbed Green's function with the following results. For a single particle in a band, the result from (2.59) is

$$G^{(0)}(\mathbf{p}, E) = \frac{1}{E - \varepsilon_{\mathbf{p}} + i\delta}$$

For a fermion in a degenerate electron gas, the result (2.64) is

$$G^{(0)}(\mathbf{p}, E) = \frac{1}{E - \varepsilon_{\mathbf{p}} + i\delta_{\mathbf{p}}}$$

The Fourier transform in time is applied to each term in the S-matrix summation:

$$G(\mathbf{p}, E) = -i \sum_{n=0}^{\infty} (-i)^n \int_{-\infty}^{\infty} dt e^{iE(t-t')} \int_{-\infty}^{\infty} dt_1 \cdots \int_{-\infty}^{\infty} dt_n$$
$$\times {}_0\langle | T\hat{C}_{\mathbf{p}\sigma}(t) \hat{C}^{\dagger}_{\mathbf{p}\sigma}(t') \hat{V}(t_1) \cdots \hat{V}(t_n) | \rangle_0 \text{ (different connected)} \tag{2.106}$$

To see what sort of terms develop, consider the example of the electron–phonon interaction. The first two terms [Figs. 2.2(c) and 2.2(e) are zero] are $G^{(0)}$ plus the self-energy term in Fig. 2.2(a):

$$G(\mathbf{p}, E) = G^{(0)}(\mathbf{p}, E) + (-i)^2 \sum_{\mathbf{q}} |M_{\mathbf{q}}|^2 \int_{-\infty}^{\infty} dt e^{iE(t-t')} \int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2$$
$$\times G^{(0)}(\mathbf{p}, t - t_1) G^{(0)}(\mathbf{p} - \mathbf{q}, t_1 - t_2)$$
$$\times G^{(0)}(\mathbf{p}, t_2 - t') D^{(0)}(\mathbf{q}, t_1 - t_2) \tag{2.107}$$

The phonon Green's function of energy is defined the same way:

$$D(\mathbf{q}, \omega) = \int_{-\infty}^{\infty} dt e^{it\omega} D(\mathbf{q}, t)$$
$$D(\mathbf{q}, t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-it\omega} D(\mathbf{q}, \omega) \tag{2.108}$$

Using the unperturbed phonon Green's function in (2.107),

$$D^{(0)}(\mathbf{q}, t_1 - t_2) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega(t_1 - t_2)} D^{(0)}(\mathbf{q}, \omega) \tag{2.109}$$

and the remaining time integrals are easy:

$$\int_{-\infty}^{\infty} dt e^{i(t-t_1)E} G^{(0)}(\mathbf{p}, t-t_1) \int_{-\infty}^{\infty} dt_1 e^{i(t_1-t_2)(E-\omega)} G^{(0)}(\mathbf{p}-\mathbf{q}, t_1-t_2)$$
$$\times \int_{-\infty}^{\infty} dt_2 e^{i(t_2-t')E} G^{(0)}(\mathbf{p}, t_2-t') = G^{(0)}(\mathbf{p}, E)^2 G^{(0)}(\mathbf{p}-\mathbf{q}, E-\omega)$$

The first two terms are

$$G(\mathbf{p}, E) = G^{(0)}(\mathbf{p}, E) + G^{(0)}(\mathbf{p}, E)^2 \Sigma^{(1)}(\mathbf{p}, E) + \cdots \tag{2.110}$$

where the self-energy of the electron due to one-phonon processes is

$$\Sigma^{(1)}(\mathbf{p}, E) = i \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \sum_{\mathbf{q}} |M_{\mathbf{q}}|^2 D^{(0)}(\mathbf{q}, \omega) G^{(0)}(\mathbf{p}-\mathbf{q}, E-\omega) \tag{2.111}$$

This self-energy will be evaluated in Chapter 7.

The electron–phonon interaction has four connected diagrams in $n = 4$; they are self-energy diagrams with two phonons and are shown in Fig. 2.7. These four terms give, respectively, the contribution to the Green's function series (2.104):

$$\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} dt_3 \int_{-\infty}^{\infty} dt_4 G^{(0)}(\mathbf{p}, t-t_1) \Bigg(\sum_{\mathbf{q}\mathbf{q}'} |M_{\mathbf{q}} M_{\mathbf{q}'}|^2 D^{(0)}(\mathbf{q}, t_1-t_2)$$
$$\times D^{(0)}(\mathbf{q}', t_3-t_4) \{ [G^{(0)}(\mathbf{p}+\mathbf{q}, t_1-t_2) G^{(0)}(\mathbf{p}, t_2-t_3) G^{(0)}(\mathbf{p}+\mathbf{q}', t_3-t_4)$$
$$+ G^{(0)}(\mathbf{p}+\mathbf{q}, t_1-t_3) G^{(0)}(\mathbf{p}+\mathbf{q}+\mathbf{q}', t_3-t_2) G^{(0)}(\mathbf{p}+\mathbf{q}', t_2-t_4)] G^{(0)}(\mathbf{p}, t_4-t')$$
$$+ G^{(0)}(\mathbf{p}+\mathbf{q}, t_1-t_3) G^{(0)}(\mathbf{p}+\mathbf{q}+\mathbf{q}', t_3-t_4) G^{(0)}(\mathbf{p}+\mathbf{q}, t_4-t_2) G^{(0)}(\mathbf{p}, t_2-t') \}$$
$$+ \sum_{\mathbf{q}} |M_{\mathbf{q}}|^4 D^{(0)}(\mathbf{q}, t_1-t_2) D^{(0)}(\mathbf{q}, t_3-t_4) G^{(0)}(\mathbf{p}+\mathbf{q}, t_1-t_4) G^{(0)}(\mathbf{p}, t_4-t')$$
$$\times \sum_{\mathbf{k},\sigma} G^{(0)}(\mathbf{k}, t_2-t_3) G^{(0)}(\mathbf{k}+\mathbf{q}, t_3-t_2) \Bigg)$$

Figure 2.7 shows the labeling of electron and phonon Green's functions and the time label of each vertex. The Fourier transform of these terms is taken to give their contribution to the Green's function of energy,

$$G^{(0)}(\mathbf{p}, E)^3 \Sigma^{(1)}(\mathbf{p}, E)^2 + G^{(0)}(\mathbf{p}, E)[\Sigma^{(2a)}(\mathbf{p}, E) + \Sigma^{(2b)}(\mathbf{p}, E) + \Sigma^{(2c)}(\mathbf{p}, E)]$$

so that combining this result with the earlier one in (2.110) gives

$$G(\mathbf{p}, E) = G^{(0)}(\mathbf{p}, E)\{1 + G^{(0)}[\Sigma^{(1)} + \Sigma^{(2a)} + \Sigma^{(2b)} + \Sigma^{(2c)}] + [G^{(0)}\Sigma^{(1)}]^2 \ldots\} \tag{2.112}$$

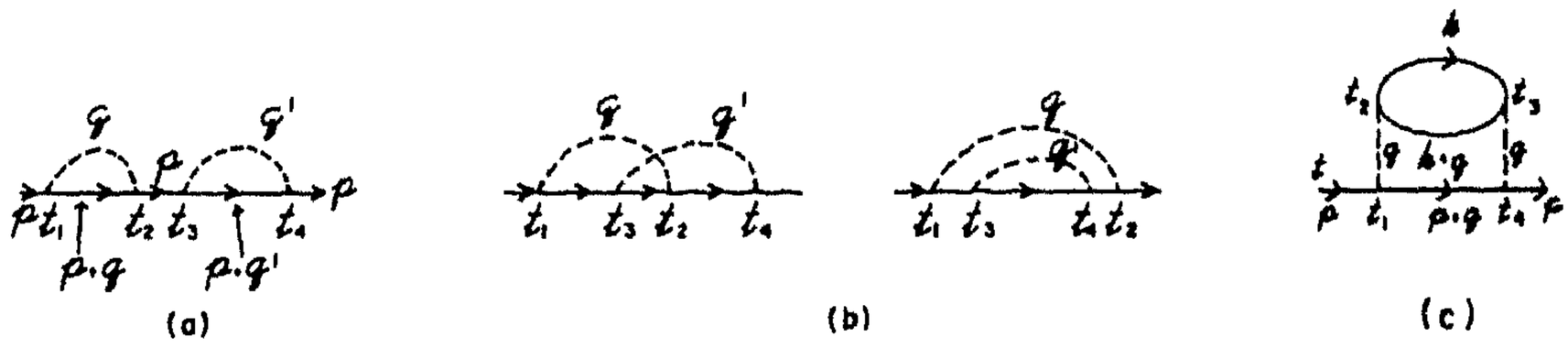


FIGURE 2.7 Two-phonon Feynman diagrams in the self-energy of the electron.

The first term, which comes from Fig. 2.7(a), contains three unperturbed Green's functions $G^{(0)}$ and two one-phonon self-energy terms $\Sigma^{(1)}$. Each of these factors can be associated with a piece of Fig. 2.7(a), as illustrated in Fig. 2.8. Similarly, the last three terms in Fig. 2.7 contain the three diagrams which represent the two-phonon self-energy terms:

$$\begin{aligned}
&\Sigma^{(2a)}(\mathbf{p}, E) + \Sigma^{(2b)}(\mathbf{p}, E) \\
&= \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \sum_{\mathbf{q}\mathbf{q}'} |M_{\mathbf{q}} M_{\mathbf{q}'}|^2 D^{(0)}(\mathbf{q}, \omega) D^{(0)}(\mathbf{q}', \omega') \\
&\quad \times G^{(0)}(\mathbf{p} + \mathbf{q}, E + \omega) G^{(0)}(\mathbf{p} + \mathbf{q} + \mathbf{q}', E + \omega + \omega') \\
&\quad \times [G^{(0)}(\mathbf{p} + \mathbf{q}', E + \omega') + G^{(0)}(\mathbf{p} + \mathbf{q}, E + \omega)] \qquad (2.113) \\
&\Sigma^{(2c)}(\mathbf{p}, E) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \sum_{\mathbf{q}} |M_{\mathbf{q}}|^4 D^{(0)}(\mathbf{q}, \omega)^2 G^{(0)}(\mathbf{p} + \mathbf{q}, E + \omega) \\
&\quad \times \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \sum_{\mathbf{k}\sigma} G^{(0)}(\mathbf{k}, \omega') G^{(0)}(\mathbf{k} + \mathbf{q}, \omega + \omega')
\end{aligned}$$

Dyson's equation is obtained by formally summing the series in (2.112),

$$G(\mathbf{p}, E) = \frac{G^{(0)}(\mathbf{p}, E)}{1 - G^{(0)}(\mathbf{p}, E)\Sigma(\mathbf{p}, E)} \qquad (2.114)$$

where the total self-energy $\Sigma(\mathbf{p}, E)$ is the summation of all different self-energy contributions:

$$\Sigma(\mathbf{p}, E) = \sum_j \Sigma^{(j)}(\mathbf{p}, E) \qquad (2.115)$$

So far the example of the electron–phonon interaction has yielded the following four terms for the self-energy:

$$\Sigma(\mathbf{p}, E) = \Sigma^{(1)}(\mathbf{p}, E) + \Sigma^{(2a)}(\mathbf{p}, E) + \Sigma^{(2b)}(\mathbf{p}, E) + \Sigma^{(2c)}(\mathbf{p}, E) + \cdots$$

Contributions such as Fig. 2.7(a) which contain $(\Sigma^{(1)})^2$ do not mean that Σ contains $(\Sigma^{(1)})^2$. The Green's function expansion contains terms in $(\Sigma^{(1)})^2$ because the expansion for G is

$$G = \frac{G^{(0)}}{1 - G^{(0)}\Sigma} = G^{(0)} + G^{(0)2}\Sigma + G^{(0)3}\Sigma^2 + G^{(0)4}\Sigma^3 + \cdots$$

Successive terms contain higher powers of each $\Sigma^{(j)}$ plus cross terms. Dyson's equation is really a theorem which states that one may sum the series of self-energy terms which develops in higher order, and the form (2.114) is obtained with each distinct contribution to Σ occurring just once.

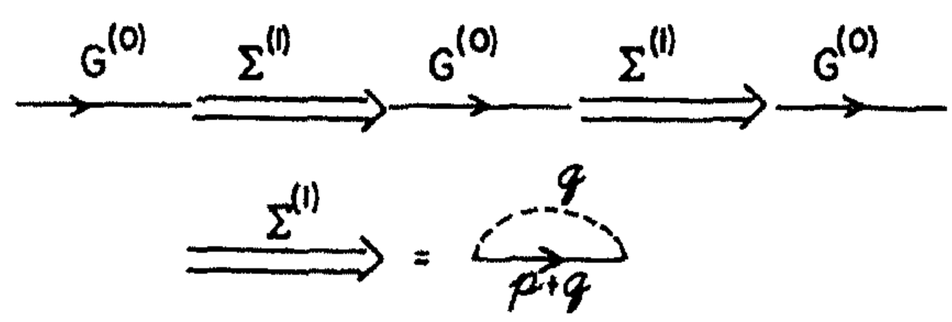


FIGURE 2.8 These four terms give, respectively, the contribution to the Green's function series (2.104):

The derivation of Dyson's equation has been rather complicated and consisted of a large number of steps and theorems. Each step has to be understood before the final result is understood. Yet the final result—Dyson's equation—has achieved a great simplification. It states that the exact Green's function is obtained from (2.114) by just calculating the self-energy $\Sigma(\mathbf{p}, E)$. The self-energy is a summation of an infinite number of distinct diagrams. This method is useful only if $\Sigma(\mathbf{p}, E)$ can be approximated by the lowest few terms in the series. Alternately, sometimes one can sum subsets of diagrams in the series. However, except in a few rare cases, it is impossible to get $\Sigma(\mathbf{p}, E)$ exactly, and one must be content with an approximate result. If the approximate result is not a very good approximation, one should not try to solve the problem in this fashion. Some of the alternate methods will be developed in subsequent chapters. But basically one should realize that Dyson's equation is usually useful only in weak coupling theory where the perturbation is sufficiently weak that an adequate approximation is obtained with a few terms in $\Sigma(\mathbf{p}, E)$. Later chapters discuss some strong coupling theories which use Dyson's equation.

A great simplification has been achieved. It is only necessary to evaluate a few self-energy diagrams in $\Sigma(\mathbf{p}, E)$. If they are sufficient, the calculation is finished. If they are not, then discard the results and try something else. The formidable-looking series (2.76), which served as the starting point, has been reduced to the evaluation of a few terms.

Dyson's equation is often written in a slightly different but equivalent form. It is obtained by using the algebraic form for $G^{(0)}$. For one electron in a band,

$$G^{(0)}(\mathbf{p}, E) = \frac{1}{E - \varepsilon_{\mathbf{p}} + i\delta}$$

$$G(\mathbf{p}, E) = \frac{1}{E - \varepsilon_{\mathbf{p}} + i\delta - \Sigma(\mathbf{p}, E)} \tag{2.116}$$

When the electron is in a Fermi sea at zero temperature,

$$G^{(0)}(\mathbf{p}, E) = \frac{1}{E - \varepsilon_{\mathbf{p}} + i\delta_{\mathbf{p}}} \tag{2.117}$$

$$G(\mathbf{p}, E) = \frac{1}{E - \varepsilon_{\mathbf{p}} + i\delta_{\mathbf{p}} - \Sigma(\mathbf{p}, E)} \tag{2.118}$$

One aspect of this result deserves special mention. The infinitesimal part $\delta_{\mathbf{p}}$ switches sign depending on whether $\varepsilon_{\mathbf{p}} > \mu$ or $\varepsilon_{\mathbf{p}} < \mu$, where μ is the chemical potential. Taking the imaginary part of (2.117) gives $E = \varepsilon_{\mathbf{p}}$. Therefore $\mathrm{Im}G^{(0)}(\mathbf{p}, E)$ changes sign depending on whether $E > \mu$ or $E < \mu$.

The self-energy has real (Σ_R) and imaginary (Σ_I) parts which are written as

$$\Sigma(\mathbf{p}, E) = \Sigma_R(\mathbf{p}, E) + i\Sigma_I(\mathbf{p}, E) \tag{2.119}$$

Later it is shown that Σ_I also switches sign at $E = \mu$ in the same manner:

$$\Sigma_I(\mathbf{p}, E) < 0, \quad E > \mu$$

$$\Sigma_I(\mathbf{p}, E) > 0, \quad E < \mu \tag{2.120}$$

The switching of signs of $\delta_{\mathbf{p}}$ was caused by the distinction between electron excitations with $\xi_{\mathbf{p}} > 0$ and hole excitations with $\xi_{\mathbf{p}} < 0$. This distinction is maintained even when $\Sigma(\mathbf{p}, E)$ is included, i.e., in the presence of interactions. The one-electron Green's function (2.116) has

the feature that $\Sigma_I < 0$, so the imaginary part of the denominator always has the same sign. The electron self-energy is sometimes called a *mass operator.*

The phonon Green's function has the same type of Dyson equation:

$$D(\mathbf{q}, \omega) = \frac{D^{(0)}(\mathbf{q}, \omega)}{1 - D^{(0)}(\mathbf{q}, \omega)\pi(\mathbf{q}, \omega)} \tag{2.121}$$

where the self-energy function for phonons is $\pi(\mathbf{q}, \omega)$. This equation may also be written in an alternate form by utilizing the following result for $D^{(0)}$:

$$D^{(0)}(\mathbf{q}, \omega) = \frac{2\omega_{\mathbf{q}}}{\omega^2 - \omega_{\mathbf{q}}^2 + i\delta} \tag{2.122}$$

$$D(\mathbf{q}, \omega) = \frac{2\omega_{\mathbf{q}}}{\omega^2 - \omega_{\mathbf{q}}^2 + i\delta - 2\omega_{\mathbf{q}}\pi(\mathbf{q}, \omega)} \tag{2.123}$$

The phonon self-energy term $\pi(\mathbf{q}, \omega)$ is sometimes called a *polarization operator.* This name is quite descriptive, since the self-energy effects arise from the phonons causing polarization in the medium.

The real and imaginary parts of the self-energies Σ and π each have interpretations. The imaginary part, Σ_I or π_I, is interpreted as causing the damping of the particle motion. They are related to the mean free path of the excitation or its energy and momentum uncertainty. The real parts are actual energy shifts of the excitation, which may also change its dynamical motion. The excitation may alter its effective mass or group velocity because of the self-energy contributions.

2.8. RULES FOR CONSTRUCTING DIAGRAMS

It is certainly worthwhile for the student to evaluate some self-energy diagrams by the method outlined above. One should expand the S matrix, decide which terms are connected and which are zero, and finally obtain the self-energies by a Fourier transform. But this laborious procedure can be avoided because the self-energy diagrams can be written down directly by following a few simple rules. That is, self-energy expressions such as (2.111) and (2.113) are easily written down in the form shown in these equations. The evaluation of the wave vector integrals in these expressions remains a formidable task. These rules are as follows:

1. Draw the Feynman diagram for the self-energy term, with all phonon, Coulomb, and electron lines.
2. For each electron line, introduce the following Green's function:

$$G_{\alpha\beta}^{(0)}(\mathbf{p}, E) = \frac{\delta_{\alpha\beta}}{E - \varepsilon_{\mathbf{p}} + i\delta_{\mathbf{p}}} \tag{2.124}$$

The $\delta_{\alpha\beta}$ conserves the spin index and indicates that the electron line must have the same spin at both ends of the propagator line. This feature is important in spin problems. The factor $\delta_{\mathbf{p}}$ is δ for one electron in a band and δ sgn $(\xi_{\mathbf{p}})$ for degenerate Fermi systems, i.e., those with a Fermi sea at zero temperature.

3. For each phonon line, introduce the following phonon propagator:

$$D^{(0)}(\mathbf{q}, \omega) = \frac{2\omega_{\mathbf{q}}}{\omega^2 - \omega_{\mathbf{q}}^2 + i\delta} \tag{2.125}$$

   Also add a factor of $|M_{\mathbf{q}}|^2$ for each phonon Green's function, where $M_{\mathbf{q}}$ is the matrix element for the electron–phonon interaction.
4. Add a Coulomb potential $v_q = 4\pi e^2/q^2$ for each Coulomb interaction. Note that the Coulomb line is always drawn as a wiggly vertical line. The Coulomb interaction is regarded as happening instantaneously in time, and time flows horizontally, from left to right, in our diagrams. One could, of course, have time flow upwards and draw the Coulomb interactions as horizontal wiggly lines.
5. Conserve energy and momentum at each vertex. Each electron line, phonon line, and Coulomb line have their variables labeled to conform with this rule.
6. Sum over internal degrees of freedom: momentum, energy, and spin. If one is calculating a self-energy term $\Sigma(\mathbf{p}, E)$ of the electron, then all momentum and energies except $\mathbf{p}$ and E are internal and must be summed over.
7. Finally, multiply the result by the factor

$$\frac{i^m}{(2\pi)^{4m}}(-1)^F(2S+1)^F \tag{2.126}$$

   where F is the number of closed fermion loops. The index m is chosen as follows:

   (a) For electron self-energies, m is the number of internal phonon and Coulomb lines. For example, $m = 1$ for (2.111), and $m = 2$ for (2.113).
   (b) For phonon self-energies, m is one-half the number of vertices. The spin of the particle is S, and the factor $(2S + 1)$ is from the summation over spin quantum number m_s. Electrons have $2S + 1 = 2$. The factor $(2\pi)^{4m}$ assumes taking the limit $\mathrm{v} \to \infty$, so that the wave vector summations are integrals. For box normalization in a finite volume v, the factor is

$$\frac{i^m}{(2\pi\mathrm{v})^m}(-1)^F(2S+1)^F \tag{2.127}$$

   and then wave vector summations are discrete summations.

The photon Green's functions are discussed in the last section. But it seems tidy to present the rules for constructing diagrams with photons at this point, so that all the rules are together. One draws photon lines between lines which represent charged particles. The photon lines are usually dotted and usually are represented just like phonon lines. Charged particles interact with the photons through two terms in the interaction Hamiltonian. The first is the $\mathbf{j} \cdot \mathbf{A}$ term. For free particles, this has the form

$$\begin{aligned}\frac{e}{c}\sum_i \mathbf{j}(\mathbf{r}_i)\cdot\mathbf{A}(\mathbf{r}_i) &= \frac{e}{c}\sum_{\mathbf{q}\mu} j_\mu(\mathbf{q})A_\mu(\mathbf{q}) \\ &= \frac{e}{mc}\sum_{\mathbf{q}\mu} A_\mu(\mathbf{q})\sum_{\mathbf{k}\sigma}(\mathbf{k}+\tfrac{1}{2}\mathbf{q})_\mu C^\dagger_{\mathbf{k}+\mathbf{q},\sigma}C_{\mathbf{k}\sigma}\end{aligned} \tag{2.128}$$

This interaction brings us to the next rule for constructing diagrams.

8. For each photon line which interacts with particles through the $\mathbf{j} \cdot \mathbf{A}$ interaction, insert a factor

$$\frac{e^2}{m^2} \sum_{\mu\nu} (\mathbf{k} + \tfrac{1}{2}\mathbf{q})_\mu D_{\mu\nu}(\mathbf{q}, \omega)(\mathbf{k}' + \tfrac{1}{2}\mathbf{q})_\nu \tag{2.129}$$

where $D_{\mu\nu}(q, \omega)$ is the photon Green's function and $\mathbf{k}$ and $\mathbf{k}'$ are the wave vectors of particles scattered at the two vertices. The other possible interaction of a charged particle with photons occurs through the term

$$\frac{e^2}{2mc^2} \sum_i A(\mathbf{r}_i)^2 = \frac{e^2}{2m} \sum_{\mathbf{q}\mathbf{k}\mu} \rho(\mathbf{q}) A_\mu(\mathbf{k}) A_\mu(\mathbf{q} - \mathbf{k})$$

In Sec. 4.5 it is shown that this interaction contributes a self-energy term of $e^2 n_0/m$ to the self-energy of the photon, where n_0 is the density of charged particles.

Other texts often define the phonon Green's function differently than the form which is selected here. However, all of these cases have the final product of the vertex and Green's function as

$$\frac{2\omega_{\mathbf{q}} |M_{\mathbf{q}}|^2}{\omega^2 - \omega_{\mathbf{q}}^2 + i\delta} \tag{2.130}$$

For example, one choice has a Green's function of the form

$$D^{(0)}(\mathbf{q}, \omega) = \frac{\omega_{\mathbf{q}}^2}{\omega^2 - \omega_{\mathbf{q}}^2 + i\delta} \tag{2.131}$$

Then, in the rules for constructing diagrams, one multiplies by the factor $2|M_{\mathbf{q}}|^2/\omega_{\mathbf{q}}$ for each vertex pair. Of course, these steps give the same result.

Now examine some examples. Electron–phonon examples are given previously in (2.111) and (2.113). Some phonon self-energies are shown in Fig. 2.9(a). The first example is from the electron–phonon interaction, and the self-energy contribution is a closed fermion loop, so that $F = 1$ in rule 7. This contribution is

$$\pi(\mathbf{q}, \omega) = |M_{\mathbf{q}}|^2 P^{(1)}(\mathbf{q}, \omega)$$

$$P^{(1)}(\mathbf{q}, \omega) = \frac{-2i}{(2\pi)^4} \int dE \int d^3p G^{(0)}(\mathbf{p}, E) G^{(0)}(\mathbf{p} + \mathbf{q}, E + \omega) \tag{2.132}$$

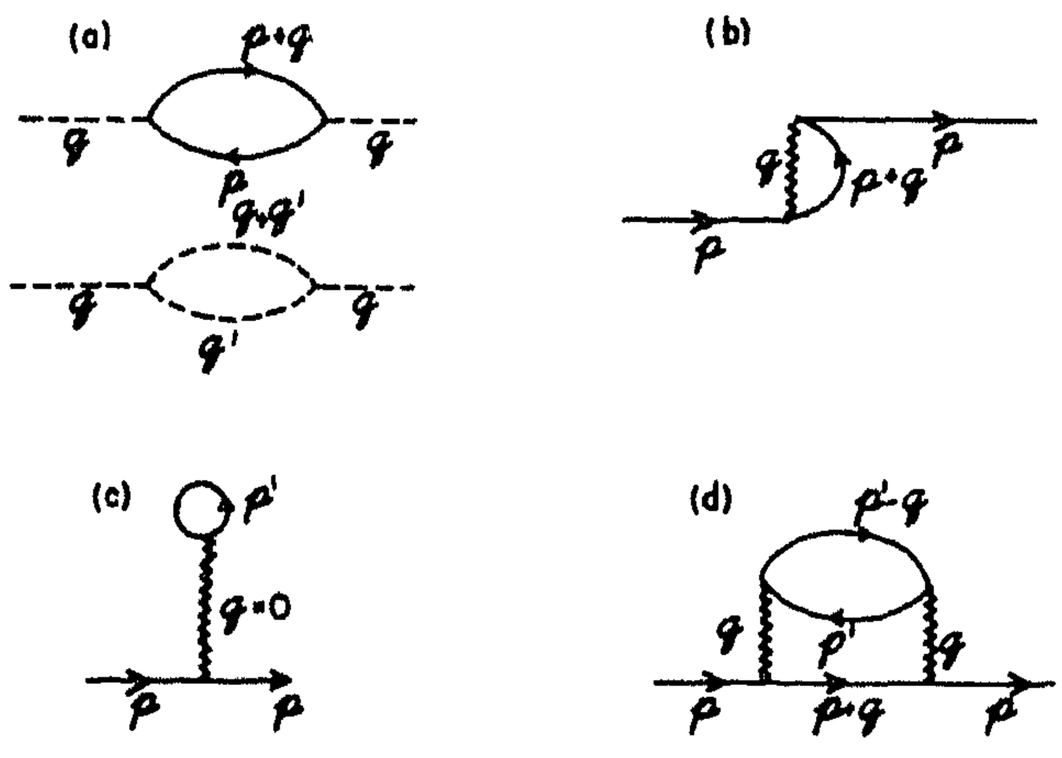


FIGURE 2.9

The other phonon self-energy contribution arises from the lattice anharmonicity, which leads to interaction terms in the ion Hamiltonian proportional to the third power of the phonon displacement:

$$V = \sum_i V_i Q_i^3 = \sum_{\mathbf{q}\mathbf{q}'} M_{\mathbf{q},\mathbf{q}'} A_{\mathbf{q}} A_{\mathbf{q}'} A_{-\mathbf{q}-\mathbf{q}'}$$

Here the self-energy term is

$$\pi^{(1)}(\mathbf{q}, \omega) = \frac{i}{(2\pi)^4} \int d\omega' \int d^3q' |M_{\mathbf{q},\mathbf{q}'}|^2 D^{(0)}(\mathbf{q}', \omega') D^{(0)}(\mathbf{q}+\mathbf{q}', \omega+\omega') \tag{2.133}$$

Next consider terms in the electron self-energy arising from electron–electron interactions. The self-energy term in Fig. 2.9(b) is called the *unscreened exchange energy* and is a very important contribution to the electron's energy. Its self-energy is

$$\Sigma_x(\mathbf{p}, E) = \frac{i}{(2\pi)^4} \int d\omega \int d^3q v_q G^{(0)}(\mathbf{p}+\mathbf{q}, E+\omega) \tag{2.134}$$

This result can be simplified immediately using an important identity:

$$i \int \frac{d\omega}{2\pi} G^{(0)}(\mathbf{p}+\mathbf{q}, E+\omega) = -n_F(\xi_{\mathbf{p}+\mathbf{q}}) \tag{2.135}$$

The identity is proved by replacing the Green's function by its Fourier definition,

$$i \int \frac{d\omega}{2\pi} \int_{-\infty}^{\infty} dt e^{it(E+\omega)} G^{(0)}(\mathbf{p}+\mathbf{q}, t) \tag{2.136}$$

and then inverting the order of integrations. The frequency integral gives a delta function in time,

$$\int \frac{d\omega}{2\pi} e^{it(E+\omega)} = \delta(t) \tag{2.137}$$

and

$$i \int \frac{d\omega}{2\pi} G^{(0)}(\mathbf{p}+\mathbf{q}, E+\omega) = iG^{(0)}(\mathbf{p}+\mathbf{q}, t=0)$$

The right-hand side is ambiguous. Since the Green's function is a time-ordered product, different results are obtained depending upon whether $t = 0$ is approached from plus time or negative time,

$$G^{(0)}(\mathbf{p}+\mathbf{q}, t \to 0^+) = -i\langle C_{\mathbf{p}+\mathbf{q}} C^\dagger_{\mathbf{p}+\mathbf{q}} \rangle = -i[1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})]$$

$$G^{(0)}(\mathbf{p}+\mathbf{q}, t \to 0^-) = -i\langle C^\dagger_{\mathbf{p}+\mathbf{q}} C_{\mathbf{p}+\mathbf{q}} \rangle = in_F(\xi_{\mathbf{p}+\mathbf{q}})$$

In the identity (2.135), the $t = 0^-$ result is chosen. It was stated in Sec. 2.5 that equal time operators were to be taken in this order. These steps prove the identity (2.135). The self-energy (2.134) may be written in the simple form

$$\Sigma_x(\mathbf{p}) = -\frac{1}{\mathrm{v}} \sum_{\mathbf{q}} v_q n_F(\xi_{\mathbf{p}+\mathbf{q}}) = -\int \frac{d^3q}{(2\pi)^3} v_q n_F(\xi_{\mathbf{p}+\mathbf{q}}) \tag{2.138}$$

where the subscript x denotes exchange. The self-energy is no longer a function of the energy E of the particle, since all E dependence has vanished from the right-hand side. This exchange energy is easily evaluated at zero temperature (see Problem 6 at the end of the chapter).

The next electron self-energy from electron–electron interactions is shown in Fig. 2.9(c):

$$\Sigma_H(\mathbf{p}, E) = \frac{-2i}{(2\pi)^4} \int dE' \int d^3p' v_{q=0} G^{(0)}(\mathbf{p}', E') \tag{2.139}$$

where the subscript "H" denotes "Hartree". The self-energy depends on neither $\mathbf{p}$ nor E. The identity (2.135) may be used again to produce the result

$$\Sigma_H = 2v_{q=0} \sum_{\mathbf{p}'} n_F(\xi_{\mathbf{p}'}) = v_{q=0} N_e \tag{2.140}$$

where N_e is the number of electrons. Of course, when $v_q = 4\pi e^2/q^2$, then the limit $\mathbf{q} \to 0$ gives $v_{q=0}$ equal to infinity. This term is the unscreened Coulomb energy from one electron interacting with all the other electrons in our system. This potential energy is truly a large number, which becomes infinity in the limit of an infinite system. But there must be an equal amount of positive charge in the system, and the electron interaction with the positive charge yields another large number which cancels the present divergence. The Hartree energy is defined as the net interaction energy of an electron from both of the negative and positive charge sources. It is zero in the jellium model of a metal, but is nonzero for actual systems composed of ions and conduction electrons.

The third electron self-energy diagram is shown in Fig. 2.9(d). Its evaluation yields

$$\Sigma(\mathbf{p}, E) = \frac{i}{(2\pi)^4} \int d\omega \int d^3q v_q^2 P^{(1)}(\mathbf{q}, \omega) G^{(0)}(\mathbf{p} + \mathbf{q}, E + \omega) \tag{2.141}$$

The factor $P^{(1)}(\mathbf{q}, \omega)$ is given in (2.132). It corresponds to the closed fermion loop. This polarization diagram occurs frequently, is very important, and will be evaluated in Chapter 5.

Finally, this section is ended with a short lecture on the way to draw diagrams. The types of Feynman diagrams are very intimately connected with the types of terms in the Hamiltonian. Figures 2.10(a) and 2.10(b) show the two types of electron–phonon vertices and the terms they correspond with in the Hamiltonian. In each case, an electron is destroyed in state $\mathbf{k}$ (incoming arrow) and created in $\mathbf{k} + \mathbf{q}$ (outgoing arrow). All of the electron–phonon diagrams involve only diagrams which can be constructed from these basic building blocks. Diagrams are not drawn in which one phonon makes two electrons, as shown in Fig. 2.9(c), since such terms do not occur in the Hamiltonian. Later such terms are encountered for other problems. Also, terms are not included in which two phonons are emitted while scattering one electron. These are anharmonic terms in the electron–phonon interaction. However, it has recently been suggested that they are important in some materials, so that in fact people do draw diagrams using them. The important point is that one must start with a Hamiltonian, decide which kind of vertex processes are permitted by each interaction term, and draw Feynman diagrams using only these basic building blocks.

2.9. TIME-LOOP *S* MATRIX

The S matrix is defined in Sec. 2.2. The time in (2.41) is taken over the interval $(-\infty, \infty)$. The state at $t = -\infty$ is well defined as the ground state of the noninteracting system. The interactions are turned on slowly. At $t \sim 0$ the fully interacting ground state is

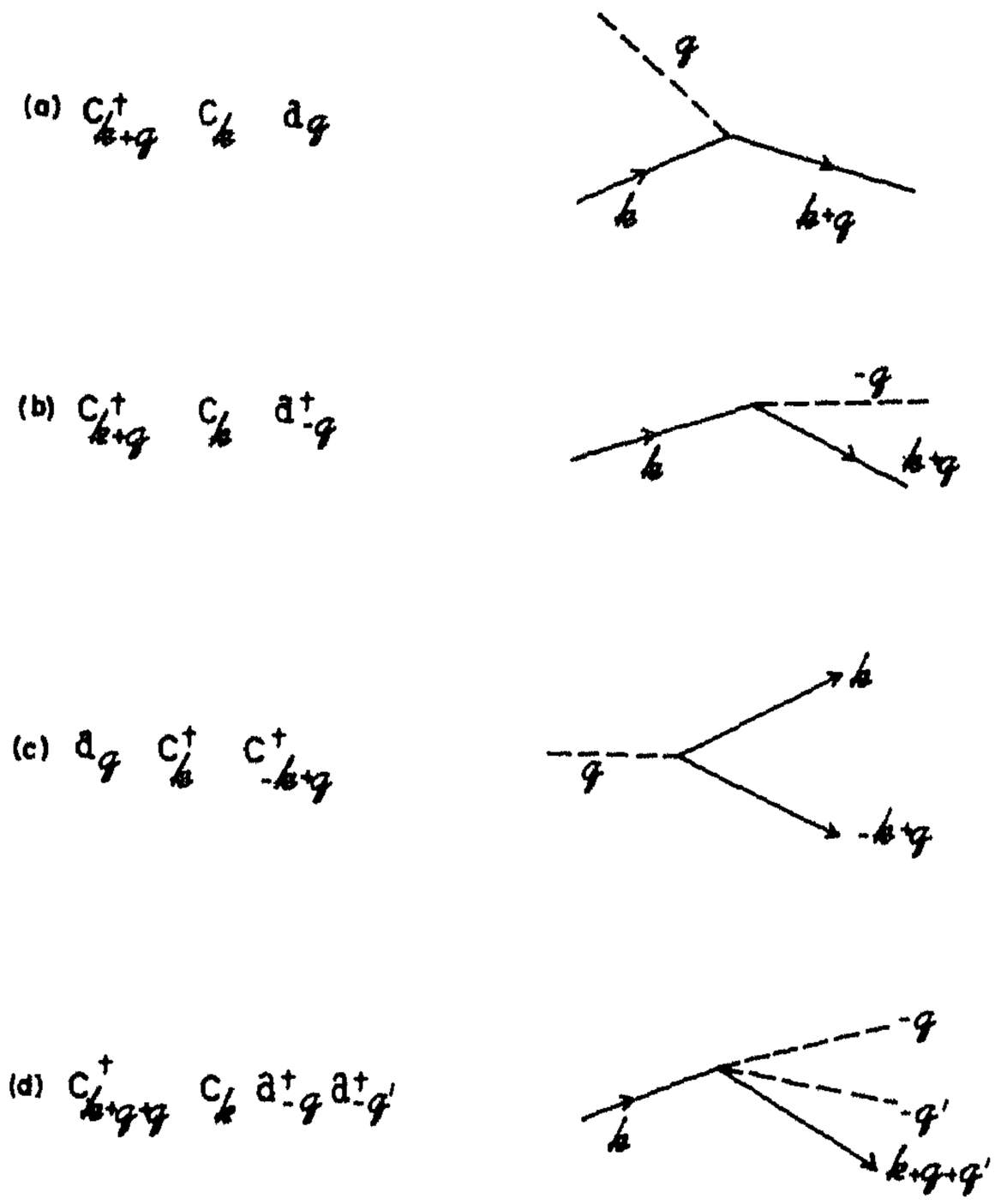


FIGURE 2.10

$\psi(0) = S(0, -\infty)\phi_0$. In condensed matter physics the state at $t \to \infty$ must be defined carefully. If the interactions remain on, then this state is not well described by the non-interacting ground state ϕ_0. Alternately, one could require that the interactions turn off at large times, which returns the system to the ground state ϕ_0.

Schwinger (1961) suggested another method of handling the asymptotic limit $t \to \infty$. He proposed that the time integral in the S matrix has two pieces: one goes from $(-\infty, \tau)$ while the second goes from $(\tau, -\infty)$. Eventually $\tau \to \infty$. The integration path is a time loop, which starts and ends at $t = -\infty$. The advantage of this method is that one starts and ends the S-matrix expansion with a known state $\psi(-\infty) = \phi_0$. Usually it is the only ground state one knows exactly.

For equilibrium phenomena the time loop method of evaluating the S matrix gives results that are identical to the other methods such as those described earlier in this chapter. A small advantage of the time-loop is that the formalism has a sounder philosophical basis since the state $\psi(\infty)$ is avoided. However, the main advantage of the time-loop method is in describing nonequilibrium phenomena using Green's functions. Nonequilibrium theory is entirely based upon this formalism, or equivalent methods. The equation of motion for the Green's function can be cast into the form of a quantum Boltzmann equation for transport theory. This application will be developed in Chapter 8.

A disadvantage of the time-loop method is that it employs six different Green's functions. They are discussed next, as a preliminary to the expansion of the S matrix. Two of these Green's functions, called retarded and advanced, are also needed for the later discussion of equilibrium theory at nonzero temperature.

The need for six Green's functions is explained by considering the time-loop expression for the S matrix:

$$S(-\infty, -\infty) = T_s \exp\left[-i \int_{\text{loop}} ds_1 \hat{V}(s_1)\right] \tag{2.142}$$

The integration path is the time loop shown in Fig. 2.11. The variable s_1 goes $(-\infty, \tau)$ and then $(\tau_1 - \infty)$. The operator T_s orders along the entire loop, with earliest values of s_1 occurring first. In expanding the S matrix, Green's functions are encountered of the form

$$G(\lambda, s_1 - s_2) = -i_0 \langle | T_s C_\lambda(s_1) C_\lambda^\dagger(s_2) | \rangle_0 \tag{2.143}$$

If both s_1 and s_2 are on the top, or outward leg, of the loop then s-ordering is identical to time-ordering. However, if both s_1 and s_2 are on the lower, or backward leg, of the loop then s-ordering is the opposite of time-ordering. This case is called anti-time-ordering. Another case is when s_1 and s_2 are on different legs. Then they are automatically ordered, and independently of the values of the s arguments.

Besides the time-ordered Green's function, it is convenient to also define one that is anti-time-ordered, and others that have no time-ordering. These different cases give rise to the four Green's functions. The other two are called retarded and advanced. They are linear combinations of these four.

2.9.1. Six Green's Functions

For the time-loop expansion, it is necessary to define six different Green's functions. It is possible to employ fewer than six since they are not independent, but using six simplifies the notation. They are all correlation functions which relate the field operator $\psi(x_1)$ of the particle at one point $x_1 - (\mathbf{r}_1, t_1)$ in spacetime to the conjugate field operator $\psi^\dagger(x_2)$ at another point $x_2 = (\mathbf{r}_2, t_2)$. The six functions are the advanced G_{adv}, retarded G_{ret}, time-ordered G_t, anti-time-ordered $G_{\bar{t}}$ and $G^<$, $G^>$ which have no name:

$$\begin{aligned}
G^>(x_1, x_2) &= -i\langle \psi(x_1)\psi^\dagger(x_2)\rangle \\
G^<(x_1, x_2) &= i\langle \psi^\dagger(x_2)\psi(x_1)\rangle \\
G_t(x_1, x_2) &= \Theta(t_1 - t_2)G^>(x_1, x_2) + \Theta(t_2 - t_1)G^<(x_1, x_2) \\
G_{\bar{t}}(x_1, x_2) &= \Theta(t_2 - t_1)G^>(x_1, x_2) + \Theta(t_1 - t_2)G^<(x_1, x_2) \\
G_{\text{ret}}(x_1, x_2) &= G_t - G^< = G^> - G_{\bar{t}} \\
G_{\text{adv}}(x_1, x_2) &= G_t - G^> = G^< - G_{\bar{t}}
\end{aligned} \tag{2.144}$$

The brackets $|\rangle$ and $\langle|$ have the same meaning of Sec. 2.3 as the ground state of the interacting system. The time-ordered Green's function is the same one in (2.43). Here it has been written in the field operator representation. For homogeneous systems in equilibrium, the Green's

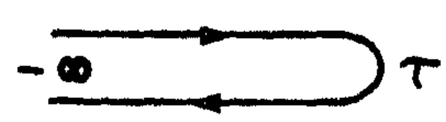


FIGURE 2.11 The time-loop integration path in the S matrix. Eventually $\tau \to \infty$.

functions depend only upon the difference of their arguments $(x_1, x_2) = (x_1 - x_2)$. Then the most simple, and useful, quantities are the Fourier transforms of these quantities:

$$G(\mathbf{k}, E) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \int dt e^{iEt} G(\mathbf{r}, t) \tag{2.145}$$

where the symbol G represents any of the six functions. Explicit expressions for these quantities are given below.

Often the Hamiltonian H can be solved exactly in terms of eigenfunctions $\phi_\lambda(\mathbf{r}_1)$ and eigenvalues ε_λ. Two examples are the electron in a magnetic field or a free particle. Then it is useful to have the expressions for the Green's functions in terms of these eigenfunctions. They are derived by expanding the field operators in terms of the eigenfunctions and creation $C_\lambda^\dagger$ and destruction C_λ operators:

$$\begin{aligned} \phi(x_1) &= \sum_\lambda C_\lambda \phi_\lambda(\mathbf{r}_1) e^{-i\varepsilon_\lambda t} \\ \phi^\dagger(x_1) &= \sum_\lambda C_\lambda^\dagger \phi_\lambda^*(\mathbf{r}_1) e^{i\varepsilon_\lambda t} \end{aligned} \tag{2.146}$$

The Green's functions in (2.144) are evaluated with the occupation factor $n_\lambda = \langle C_\lambda^\dagger C_\lambda \rangle$ and $t = t_1 - t_2$. At zero temperature $n_\lambda = \Theta(-\xi_\lambda)$ is a step function that is zero or one depending upon whether $\xi_\lambda = \varepsilon_\lambda - \mu$ is positive or negative:

$$\begin{aligned} G^>(x_1, x_2) &= -i \sum_\lambda (1 - n_\lambda) \phi_\lambda(\mathbf{r}_1) \phi_\lambda^*(\mathbf{r}_2) e^{-i\varepsilon_\lambda t} \\ G^<(x_1, x_2) &= i \sum_\lambda n_\lambda \phi_\lambda(\mathbf{r}_1) \phi_\lambda^*(\mathbf{r}_2) e^{-i\varepsilon_\lambda t} \\ G_t(x_1, x_2) &= -i \sum_\lambda [\Theta(t) - n_\lambda] \phi_\lambda(\mathbf{r}_1) \phi_\lambda^*(\mathbf{r}_2) e^{-i\varepsilon_\lambda t} \\ G_{\bar{t}}(x_1, x_2) &= -i \sum_\lambda [\Theta(-t) - n_\lambda] \phi_\lambda(\mathbf{r}_1) \phi_\lambda^*(\mathbf{r}_2) e^{-i\varepsilon_\lambda t} \\ G_{\text{ret}}(x_1, x_2) &= -i\Theta(t) \sum_\lambda \phi_\lambda(\mathbf{r}_1) \phi_\lambda^*(\mathbf{r}_2) e^{-i\varepsilon_\lambda t} \\ G_{\text{adv}}(x_1, x_2) &= i\Theta(-t) \sum_\lambda \phi_\lambda(\mathbf{r}_1) \phi_\lambda^*(\mathbf{r}_2) e^{-i\varepsilon_\lambda t} \end{aligned} \tag{2.147}$$

The above formulas are also valid in equilibrium at nonzero temperatures if $n_\lambda = 1/[\exp(\beta\xi_\lambda) + 1]$ is the thermodynamic average ($\beta = 1/k_B T$) of the occupation number.

The starting point for any calculation, at least conceptually, is the behavior of the Green's functions for systems without interactions. Then the wave functions are those for plane wave or noninteracting Bloch states, if such can be defined. The quantum number λ becomes the wave vector $\mathbf{k}$, and a spin index σ, which is usually not written. The eigenvalue combination is $\phi_\lambda(\mathbf{r}_1)\phi_\lambda^*(\mathbf{r}_2) = \exp[i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)]/\nu$. The superscript "(0)" or the subscript "0" on the Green's functions means those for a noninteracting system in equilibrium. Fourier-trans-

forming the $\mathbf{r}$ variable to $\mathbf{k}$ as in (2.145) gives the free-particle Green's functions $G^{(0)}(\mathbf{k}, t)$. For fermions of band energy $\varepsilon_{\mathbf{k}}$ and occupation number $n_{\mathbf{k}} = n_F(\varepsilon_{\mathbf{k}} - \mu)$ they are

$$
\begin{aligned}
G_0^{>}(\mathbf{k}, t) &= -i[1 - n_{\mathbf{k}}]e^{-i\varepsilon_{\mathbf{k}}t} \\
G_0^{<}(\mathbf{k}, t) &= in_{\mathbf{k}}e^{-i\varepsilon_{\mathbf{k}}t} \\
G_t^{(0)}(\mathbf{k}, t) &= -i[\Theta(t) - n_{\mathbf{k}}]e^{-i\varepsilon_{\mathbf{k}}t} \\
G_{\bar{t}}^{(0)}(\mathbf{k}, t) &= -i[\Theta(-t) - n_{\mathbf{k}}]e^{-i\varepsilon_{\mathbf{k}}t} \\
G_{\text{ret}}^{(0)}(\mathbf{k}, t) &= -i\Theta(t)e^{-i\varepsilon_{\mathbf{k}}t} \\
G_{\text{adv}}^{(0)}(\mathbf{k}, t) &= i\Theta(-t)e^{-i\varepsilon_{\mathbf{k}}t}
\end{aligned}
\tag{2.148}
$$

The t variable can be Fourier transformed, which gives the noninteracting Green's function of energy E; the quantity δ is infinitesimal

$$
\begin{aligned}
G_0^{>}(\mathbf{k}, E) &= -2\pi i[1 - n_{\mathbf{k}}]\delta(E - \varepsilon_{\mathbf{k}}) \\
G_0^{<}(\mathbf{k}, E) &= 2\pi i n_{\mathbf{k}}\delta(E - \varepsilon_{\mathbf{k}}) \\
G_{\text{ret}}^{(0)}(\mathbf{k}, E) &= \frac{1}{E - \varepsilon_{\mathbf{k}} + i\delta} \\
G_{\text{adv}}^{(0)}(\mathbf{k}, E) &= \frac{1}{E - \varepsilon_{\mathbf{k}} - i\delta} \\
G_t^{(0)}(\mathbf{k}, E) &= G_{\text{ret}}^{(0)} + G_0^{<} = \frac{1}{E - \varepsilon_{\mathbf{k}} + i\delta_{\mathbf{k}}} \\
G_{\bar{t}}^{(0)}(\mathbf{k}, E) &= G_0^{<} - G_{\text{adv}}^{(0)} = \frac{-1}{E - \varepsilon_{\mathbf{k}} - i\delta_{\mathbf{k}}}
\end{aligned}
\tag{2.149}
$$

The time-ordered function $G_t^{(0)}$ is exactly the same one in (2.64). Note the two kinds of infinitesimal deltas: δ is always positive, while $\delta_{\mathbf{k}}$ is positive for $k > k_F$ and negative for $k < k_F$ as defined earlier in (2.64). The retarded functions always have a positive δ, even for electrons in a partially filled band. The noninteracting advanced Green's function resembles that of an empty band in (2.59). These two Green's functions could differ as soon as interactions are introduced, since they have different self-energy functions in degenerate Fermi systems. Also note that expressions such as $G_{\text{ret}} = G_t - G^{<}$ are obeyed for interacting and noninteracting functions They are obeyed for both cases of arguments $(\mathbf{k}, t)$ and $(\mathbf{k}, E)$.

The above Green's functions are suitable for particles such as electrons, or holes in semiconductors. Another type of Green's function is needed for boson fields such as phonons or photons. For phonons let $Q(x)$ be the displacement from equilibrium of the ions in the solid

at position $x = (\mathbf{r}, t)$ in spacetime. Displacements are vectors, but treat them as scalars for the moment. The phonon Green's functions are defined as follows:

$$
\begin{aligned}
D^{>}(x_1, x_2) &= -i\langle Q(x_1)Q(x_2)\rangle \\
D^{<}(x_1, x_2) &= -i\langle Q(x_2)Q(x_1)\rangle \\
D_t(x_1, x_2) &= \Theta(t_1 - t_2)D^{>}(x_1, x_2) + \Theta(t_2 - t_1)D^{<}(x_1, x_2) \\
D_{\bar{t}}(x_1, x_2) &= \Theta(t_2 - t_1)D^{>}(x_1, x_2) + \Theta(t_1 - t_2)D^{<}(x_1, x_2) \\
D_{\text{ret}} &= D_t - D^{<} = \Theta(t_1 - t_2)[D^{>} - D^{<}] \\
D_{\text{adv}} &= D_t - D^{>} = -\Theta(t_2 - t_1)[D^{>} - D^{<}]
\end{aligned}
\tag{2.150}
$$

These expressions are rather similar to those in (2.144) for particles. The main difference is that $D^{<}$ and $D^{>}$ have the same sign, since no sign change is made when interchanging the positions of boson operators. Also the displacement operator is Hermitian $[Q^{\dagger} = Q]$, which introduces some redundancy such as $D^{<}(x_1, x_2) = D^{>}(x_2, x_1)$.

The displacement operators Q are usually represented in terms of phonon raising $(a^{\dagger})$ and lowering (a) operators. The usual case is to use $A_{\mathbf{q}} = a^{\dagger}_{-\mathbf{q}} + a_{\mathbf{q}}$ instead of $Q(x)$ in the definition of the phonon Green's function. In this representation, the phonon Green's functions in equilibrium are expressed in terms of the phonon occupation number $N_{\mathbf{q}} = \langle a^{\dagger}_{\mathbf{q}} a_{\mathbf{q}}\rangle$, which equals $1/[\exp(\beta\omega_{\mathbf{q}}) - 1]$ in thermal equilibrium at nonzero temperature, and equals zero at $T = 0$. The noninteracting results are:

$$
\begin{aligned}
D^{>}(\mathbf{q}, t) &= -i[(N_{\mathbf{q}} + 1)e^{-i\omega_{\mathbf{q}}t} + N_{\mathbf{q}}e^{i\omega_{\mathbf{q}}t}] \\
D^{<}(\mathbf{q}, t) &= -i[(N_{\mathbf{q}} + 1)e^{i\omega_{\mathbf{q}}t} + N_{\mathbf{q}}e^{-i\omega_{\mathbf{q}}t}] \\
D_{\text{ret}}(\mathbf{q}, t) &= -2\Theta(t)\sin(\omega_{\mathbf{q}}t) \\
D_{\text{adv}}(\mathbf{q}, t) &= -2\Theta(-t)\sin(\omega_{\mathbf{q}}t) \\
D_t(\mathbf{q}, t) &= -i\{(N_{\mathbf{q}} + \Theta(-t))e^{i\omega_{\mathbf{q}}t} + [N_{\mathbf{q}} + \Theta(t)]e^{-i\omega_{\mathbf{q}}t}\} \\
D_{\bar{t}}(\mathbf{q}, t) &= -i\{(N_{\mathbf{q}} + \Theta(t))e^{i\omega_{\mathbf{q}}t} + [N_{\mathbf{q}} + \Theta(-t)]e^{-i\omega_{\mathbf{q}}t}\}
\end{aligned}
\tag{2.151}
$$

These Green's functions are used in the expansion of the S matrix.

2.9.2. Dyson's Equation

Each of the six Green's functions can be evaluated for an interacting system. They can be expressed in the interaction representation in terms of the time-loop S matrix. For example, one of them is

$$
G^{<}(x_1, x_2) = i\langle S\hat{\psi}^{\dagger}(x_2)\hat{\psi}(x_1)\rangle \tag{2.152}
$$

where the S matrix is given in (2.142). This argument has no time-ordering operator since the order of the two state operators is fixed. However, the S matrix is time-ordered. The time t_1 is

on the upper loop, while t_2 is on the return loop. The Green's function in the above equation has an S-matrix expansion of the form

$$G^<(x_1, x_2) = i \sum_n \frac{(-i)^n}{n!} \int ds_1 \cdots \int ds_n \langle T\hat{V}(s_1) \cdots \hat{V}(s_n)\hat{\psi}^\dagger(x_2)\hat{\psi}(x_1)\rangle \tag{2.153}$$

where all s integrals are over the time loop. A method is needed to evaluate expressions of this type.

The potential V is composed of electron, phonon, or photon operators. The operators are paired using Wick's theorem. Each pair will have a time argument such as $G(s_i, s_j)$. If both s_i and s_j are in the top loop, the expression is just the time-ordered Green's function. If they are both in the return loop, the expression is the anti-time-ordered Green's function. If one s variable is in the top loop and the other is in the bottom loop, then the T_s operator makes this expression be either $G^<$ or else $G^>$. These relationships are shown in Fig. 2.12. The nth term in the Green's function expansion is a product of $(n+1)$ factors, where each factor is one of the four Green's functions in Fig. 2.12.

It helps to have a simple example. Below is given a potential term V of the type found for electrons scattering from impurities. The first term in the S-matrix expansion for $G^<$ with this interaction is

$$V = \sum_{\alpha\beta} M_{\alpha\beta} C_\alpha^\dagger C_\beta$$

$$\begin{aligned} G^<(\lambda, t_1 - t_2) &= G_0^<(\lambda, t_1 - t_2) \\ &\quad + \sum_{\alpha\beta} M_{\alpha\beta} \int ds_0 \langle | T C_\lambda^\dagger(t_2) C_\beta(s) | \rangle_{00} \langle | T C_\alpha^\dagger(s) C_\lambda(t_1) | \rangle_0 \end{aligned} \tag{2.154}$$

The s integral runs over the time loop. For $G^<$ remember that t_1 is on the top loop while t_2 is on the return loop. There are two possibilities. If s is on the top loop, the first Green's function in the s integral is $G^<$ and the second is time-ordered. If s is on the return loop, the first is anti-time-ordered while the second is $G^<$. The two terms are $(s = t', \tau \to \infty)$

$$\begin{aligned} G^<(\lambda, t_1 - t_2) = G_0^<(\lambda, t_1 - t_2) + M_{\lambda\lambda} \int_{-\infty}^{\infty} dt' [& G_t^{(0)}(\lambda, t_1 - t') G_0^<(\lambda, t' - t_2) \\ & - G_0^<(\lambda, t_1 - t') G_{\bar{t}}^{(0)}(\lambda, t' - t_2)] \end{aligned} \tag{2.155}$$

A sign change occurred in the last term when the direction of the time integration was changed from $(\infty, -\infty)$ to $(-\infty, \infty)$. The above expression contains only the first two terms in the S-matrix expansion, which has an infinite number of terms.

In the expansion of the S matrix, each time integral produces one set of terms for the outward s leg, and another for the return leg. The nth term in the S-matrix expansion produces 2^n arrangements. All of these terms can be managed by using a matrix formulation.

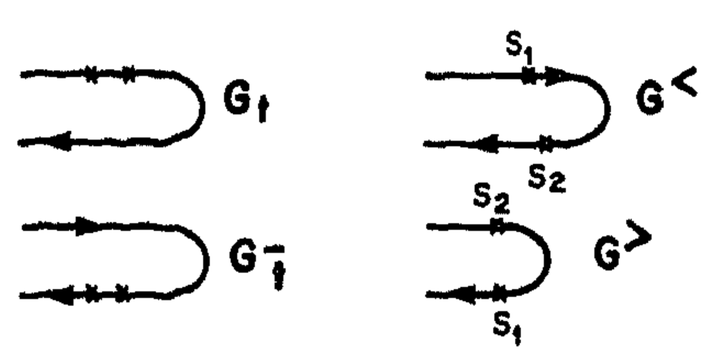


FIGURE 2.12 The four Green's functions $G(s_1, s_2)$ depend upon whether the time variables (s_1, s_2) are on the outgoing or return parts of the time loop.

Keldysh (1965) was the first to develop the time-loop theory for solid state applications. Here the theory is presented using a matrix notation suggested by Craig (1968). He expressed four of these Green's functions as the elements of a 2×2 matrix. The self-energy terms are also a matrix:

$$\tilde{G} = \begin{bmatrix} G_t & -G^< \\ G^> & -G_{\bar{t}} \end{bmatrix}$$

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma_t & -\Sigma^< \\ \Sigma^> & -\Sigma_{\bar{t}} \end{bmatrix} \tag{2.156}$$

The actual form for the self-energy functions is discussed in the following chapters.

For systems either in equilibrium or nonequilibrium, Dyson's equation is most easily expressed by using the matrix notation:

$$\tilde{G}(x_1, x_2) = \tilde{G}_0(x_1 - x_2) + \int_{-\infty}^{\infty} dx_3 \int_{-\infty}^{\infty} dx_4 \tilde{G}_0(x_1 - x_3) \times \tilde{\Sigma}(x_3, x_4)\tilde{G}(x_4, x_2) \tag{2.157}$$

The matrix formulation comes directly from the time loop. Each s integral in the S matrix has an outward and return leg. Each of these legs gives a different Green's function. So each time integral generates two Green's functions. Hence the usefulness of the 2×2 matrix formalism.

Some simple expressions can be obtained for the Green's functions. First write the above equation in a notation where the product of two functions implies an integration over the four-variable dx, which condenses the same equation to

$$\tilde{G} = \tilde{G}_0 + \tilde{G}_0\tilde{\Sigma}\tilde{G} \tag{2.158}$$

Then the equations are iterated. The following exact expressions are derived for the equations obeyed by the various Green's functions, using the same product notation:

$$\begin{aligned}
G_{\text{ret}} &= G_{\text{ret}}{}^{(0)}[1 + \Sigma_{\text{ret}} G_{\text{ret}}] \\
G_{\text{adv}} &= G_{\text{adv}}{}^{(0)}[1 + \Sigma_{\text{adv}} G_{\text{adv}}] \\
G^> &= [1 + G_{\text{ret}}\Sigma_{\text{ret}}]G_0^>[1 + \Sigma_{\text{adv}} G_{\text{adv}}] + G_{\text{ret}}\Sigma^> G_{\text{adv}} \\
G^< &= [1 + G_{\text{ret}}\Sigma_{\text{ret}}]G_0^<[1 + \Sigma_{\text{adv}} G_{\text{adv}}] + G_{\text{ret}}\Sigma^< G_{\text{adv}} \\
G_t &= [1 + G_{\text{ret}}\Sigma_{\text{ret}}]G_t^{(0)}[1 + \Sigma_{\text{adv}} G_{\text{adv}}] + G_{\text{ret}}\Sigma_{\bar{t}} G_{\text{adv}} \\
\Sigma_{\bar{t}} &= [1 + G_{\text{ret}}\Sigma_{\text{ret}}]\Sigma_{\bar{t}}^{(0)}[1 + \Sigma_{\text{adv}} G_{\text{adv}}] + G_{\text{ret}}\Sigma_t G_{\text{adv}}
\end{aligned} \tag{2.159}$$

These equations represent multiple integrals in d^3rdt.

Equations (2.159) simplify considerably for homogeneous systems in steady state where the arguments of the Green's functions and self-energies depend only upon $(x_1 - x_2)$. If the equations are Fourier transformed, then all quantities depend only upon $(\mathbf{k}, \omega)$. The above equations are, after Fourier transforming, just algebraic quantities, which are easily solved. The time-ordered Green's function has the same equation as given in (2.118). Some of the other Green's functions are given below. The expressions are presented for nonzero temperature. The zero-temperature cases are found by setting the fermion occupation factor

$n_F(\omega) = \Theta(-\omega)$. Also define a quantity $A(\mathbf{k}, \omega) = -2\mathrm{Im}G_{\mathrm{ret}}(\mathbf{k}, \omega)$ which is called the *spectral function:*

$$G_{\mathrm{ret}}(\mathbf{k}, \omega) = \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma_{\mathrm{ret}}}, \qquad \sigma = \omega - \varepsilon_{\mathbf{k}} - \mathrm{Re}\Sigma_{\mathrm{ret}}$$

$$G_{\mathrm{adv}}(\mathbf{k}, \omega) = \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma_{\mathrm{adv}}}, \qquad \Sigma_{\mathrm{adv}} = \Sigma_{\mathrm{ret}}{}^*$$

$$A(\mathbf{k}, \omega) = \frac{2\Gamma}{\sigma^2 + \Gamma^2}, \qquad \Gamma = -\mathrm{Im}\Sigma_{\mathrm{ret}} > 0$$

$$\Sigma^< = 2in_F(\omega)\Gamma(\mathbf{k}, \omega), \qquad n_F(\omega) = \frac{1}{e^{\beta\omega} + 1}$$

$$G^< = in_F(\omega)A(\mathbf{k}, \omega)$$

$$\Sigma^> = -2i(1 - n_F)\Gamma$$

$$G^> = -i(1 - n_F)A \tag{2.160}$$

The notation $\Gamma = -\mathrm{Im}\Sigma_{\mathrm{ret}}$ will be used throughout, and rather often. The spectral function is an important quantity, which is discussed further in the next chapter. These formulas are incomplete without a prescription for calculating the different self-energy functions $\Sigma^<$, Σ_{ret}, etc. They will be presented in the next chapter.

2.10. PHOTON GREEN'S FUNCTIONS

The interaction of charges with themselves and with the photon field was discussed in Sec. 1.5. For spinless particles, this interaction has the Hamiltonian in the nonrelativistic limit:

$$H = \sum_i \frac{1}{2m}\left[\mathbf{p}_i - \frac{e}{c}\mathbf{A}(\mathbf{r}_i)\right]^2 + \frac{1}{2}\sum_{i \neq j}\frac{e_i e_j}{r_{ij}} + \sum_{\mathbf{k}\lambda}\omega_{\mathbf{k}\lambda}a^\dagger_{\mathbf{k}\lambda}a_{\mathbf{k}\lambda} \tag{2.161}$$

The vector potential is given by the expansion

$$\frac{1}{c}A_\mu = \frac{1}{\sqrt{V}}\sum_{\mathbf{k}\lambda} e^{i\mathbf{k}\cdot\mathbf{r}}A_\mu(\mathbf{k}, \lambda, t) \tag{2.162}$$

$$A_\mu(\mathbf{k}, \lambda, t) = \left(\frac{2\pi}{\omega_{\mathbf{k}}}\right)^{1/2}\xi_\mu(\mathbf{k}\lambda)(a_{\mathbf{k}\lambda}e^{-i\omega\mathbf{k}^t} + a^\dagger_{-\mathbf{k}\lambda}e^{i\omega\mathbf{k}^t}) \tag{2.163}$$

The creation and destruction operators $a^\dagger_{\mathbf{k}\lambda}, a_{\mathbf{k}\lambda}$ obey boson statistics. Each state with wave vector $\mathbf{k}$ and polarization λ has its own harmonic oscillator statistics. The vector potential represents the photon field. Two charges may interact via their common photon field or more directly through the instantaneous Coulomb interaction $e_i e_j / r_{ij}$. The division of the interaction between photons and Coulomb field is arbitrary—both interactions come from the same basic processes. The Hamiltonian (2.161) is written in the Coulomb gauge where $\nabla \cdot \mathbf{A} = 0$. Another choice of gauge will result in a different division between photon and Coulomb. The basic forces between the particles are the same regardless of how the gauge is selected.

Now it is time to talk about Green's functions. The scalar potential $\psi(\mathbf{r}, t)$ has a Green's function. The potential from a point charge is

$$\psi(r) = \frac{e^2}{r} \tag{2.164}$$

The factor

$$v_q = \frac{4\pi e^2}{q^2} \tag{2.165}$$

which has already been used for the Coulomb interaction, is in fact just the Green's function of the longitudinal potential. It has no frequency dependence because it is instantaneous. After all, in the rules for constructing diagrams, in Sec. 2.8, it was treated as a Green's function. If two electrons interacted by phonons, put in the phonon Green's function and vertex,

$$|M_{\mathbf{q}}|^2 D^{(0)}(\mathbf{q}, \omega) \tag{2.166}$$

while if two electrons interact by electron–electron interactions, put in the factor v_q where $4\pi/q^2$ is regarded as the Green's function and e^2 as the vertex. The Coulomb interaction is treated as a Green's function, on equal footing with the phonon Green's function.

Since v_q is a Green's function, it has a Dyson equation of the usual form:

$$v_q(\omega) = \frac{v_q}{1 - v_q P(\mathbf{q}, \omega)} \tag{2.167}$$

The factor $P(\mathbf{q}, \omega)$ is the self-energy or polarization operator. Its properties are discussed extensively in Chapter 5. There is one result which can be obtained with very little effort. Consider the form of Maxwell's equations in a homogeneous material with an isotropic dielectric constant ε:

$$\begin{aligned} \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= 0 \\ \varepsilon \nabla \cdot \mathbf{E} &= 4\pi\rho \\ \nabla \times \mathbf{B} &= \frac{\varepsilon}{c}\frac{\partial}{\partial t}\mathbf{E} + \frac{4\pi}{c}\mathbf{j} \end{aligned} \tag{2.168}$$

If these equations are solved in the usual way, further equations are obtained for the scalar and vector potentials:

$$\begin{aligned} \psi(\mathbf{r}) &= \frac{1}{\varepsilon}\int \frac{d^3r'\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\ \nabla^2 \mathbf{A} - \frac{\varepsilon}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A} &= -\frac{4\pi}{c}\mathbf{j}_t \end{aligned} \tag{2.169}$$

Here the Coulomb Green's function is

$$\bar{v}_q = \frac{v_q}{\varepsilon} \tag{2.170}$$

If this result is regarded as equivalent to (2.167), it gives a formula for the dielectric function:

$$\varepsilon(\mathbf{q}, \omega) = 1 - v_q P(\mathbf{q}, \omega) \tag{2.171}$$

On the left-hand side this result has been generalized to include the case where the dielectric function depends on both $\mathbf{q}$ and ω. Equation (2.171) will serve as the definition of the longitudinal dielectric function. It arises from the self-energy parts of the Coulomb potential. The Green's function for the vector potential is

$$D_{\mu\nu}(\mathbf{k}, t - t') = -i \sum_{\lambda} \langle T A_\mu(\mathbf{k}, \lambda, t) A_\nu(-\mathbf{k}, \lambda, t') \rangle \tag{2.172}$$

where μ, ν are the x, y, z components. The vector potential is defined in (2.163). The sum over λ is the sum over the two transverse polarizations of the light, while ξ_μ are the polarization vectors for each component. The free propagator at zero temperature is evaluated with the states $|\rangle_0$ and ${}_0\langle|$ as the photon vacuum:

$$\begin{aligned} D^{(0)}_{\mu\nu}(\mathbf{k}, t - t') &= \frac{2\pi i}{\omega_{\mathbf{k}}} \sum_{\lambda} \xi_\mu(\mathbf{k}, \lambda) \xi_\nu(-\mathbf{k}, \lambda) {}_0\langle| T[a_{\mathbf{k}\lambda} e^{-i\omega_{\mathbf{k}} t} + a^\dagger_{-\mathbf{k}\lambda} e^{i\omega_{\mathbf{k}} t}] \\ &\quad \times [a_{-\mathbf{k}\lambda} e^{-i\omega_{\mathbf{k}} t'} + a^\dagger_{\mathbf{k}\lambda} e^{i\omega_{\mathbf{k}} t'}] |\rangle_0 \\ &= -\frac{2\pi i}{\omega \mathbf{k}} \sum_{\lambda} \xi_\mu \xi_\nu [\Theta(t' - t) e^{i\omega \mathbf{k}(t - t')} {}_0\langle| a_{-\mathbf{k}\lambda} a^\dagger_{-\mathbf{k}\lambda} |\rangle_0 \\ &\quad + \Theta(t - t') e^{i\omega \mathbf{k}(t' - t)} {}_0\langle| a_{\mathbf{k}\lambda} a^\dagger_{\mathbf{k}\lambda} |\rangle_0] \\ &= -\frac{2\pi i}{\omega_{\mathbf{k}}} e^{-i\omega_{\mathbf{k}} |t - t'|} \sum_{\lambda} \xi_\mu \xi_\nu \end{aligned} \tag{2.173}$$

and its Fourier transform is

$$D^{(0)}_{\mu\nu}(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} dt e^{i\omega(t - t')} D^{(0)}_{\mu\nu}(\mathbf{k}, t - t') \tag{2.174}$$

$$= \frac{4\pi}{\omega^2 - \omega^2_{\mathbf{k}} + i\delta} \sum_{\lambda} \xi_\mu \xi_\nu \tag{2.175}$$

Next, consider the last factor in the above equation $\sum_\lambda \xi_\mu \xi_\nu$. The unit tensor is

$$\delta_{\mu\nu} = \hat{x}\hat{x} + \hat{y}\hat{y} + \hat{z}\hat{z} \tag{2.176}$$

Or, in a coordinate system in the direction of $\mathbf{k}$,

$$\delta_{\mu\nu} = \sum_{\lambda} \xi_\mu \xi_\nu + \hat{k}\hat{k} \tag{2.177}$$

The factor $\sum_\lambda \xi_\mu \xi_\nu$ is the unit dyadic for directions perpendicular to $\mathbf{k}$. The Green's function is

$$D^{(0)}_{\mu\nu} = \frac{4\pi[\delta_{\mu\nu} - (k_\mu k_\nu / k^2)]}{\omega^2 - \omega^2_{\mathbf{k}} + i\delta} \tag{2.178}$$

This expression is referred to as the *photon Green's function*. Keep in mind that the interaction between two charges occurs via both the scalar and vector potentials. How the interaction is divided between scalar and vector potentials is somewhat arbitrary and is determined by the gauge condition. After making this choice, the word "photon" is assigned to the vector

potential part. This division between photon and Coulomb is arbitrary, and both parts should really be viewed as arising from photons.

The photon Green's function also obeys a Dyson equation. Since it is a matrix quantity, one must be careful about the treatment of indices. Previous Dyson equations have only scalar quantities, where the Green's function is defined as a series expansion. The equation for electrons is actually of the form

$$G(\mathbf{p}, E) = G^{(0)}(\mathbf{p}, E) + G^{(0)}(\mathbf{p}, E)\Sigma(\mathbf{p}, E)G(\mathbf{p}, E) \tag{2.179}$$

Since all the quantities are scalar functions, this equation could be solved immediately for G($\mathbf{p}, E$) which obtains (2.114). But for the photon Green's function, the equation is

$$D_{\mu\nu} = D^{(0)}_{\mu\nu} + \sum_{\lambda\delta} D^{(0)}_{\mu\lambda}\pi_{\lambda\delta}D_{\delta\nu} \tag{2.180}$$

where $\pi_{\lambda\delta}(\mathbf{k}, \omega)$ is the self-energy function, which is now a 3×3 matrix. Each term in the equation is a function of ($\mathbf{k}$, ω). However, in homogeneous materials all matrices are of the form

$$D^{(0)}_{\mu\nu} = \left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}\right)D^{(0)}$$

$$D_{\mu\nu} = \left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}\right)D \tag{2.181}$$

$$\pi_{\mu\nu} = \delta_{\mu\nu}\pi^{(1)} + \frac{k_\mu k_\nu}{k^2}\pi^{(2)} \tag{2.182}$$

where the factors $D^{(0)}$, D, $\pi^{(1)}$, and $\pi^{(2)}$ are scalars. The self-energy function $\pi_{\mu\nu}(\mathbf{k}, \omega)$ has the form shown, which is the most general dependence on $\mathbf{k}$ of a matrix function. Now it is simple to do the summations over the matrix components:

$$\sum_{\lambda\delta}(\delta_{\mu\lambda} - \hat{k}_\mu\hat{k}_\lambda)(\delta_{\lambda\delta}\pi^{(1)} + \hat{k}_\lambda\hat{k}_\delta\pi^{(2)})(\delta_{\delta\nu} - \hat{k}_\delta\hat{k}_\nu) = (\delta_{\mu\nu} - \hat{k}_\mu\hat{k}_\nu)\pi^{(1)}$$

and there is a scalar equation for the photon Dyson's equation

$$D = \frac{D^{(0)}}{1 - D^{(0)}\pi^{(1)}} \tag{2.183}$$

$$D_{\mu\nu} = \frac{4\pi[\delta_{\mu\nu} - k_\mu k_\nu/k^2]}{\omega^2 - \omega_{\mathbf{k}}^2 - 4\pi\pi^{(1)}} \tag{2.184}$$

The transverse photon has no dependence in its self-energy on the longitudinal part of the self-energy $\pi^{(2)}$. The equation for the photon Green's function is (2.184).

Actual solids are periodic, rather than homogeneous, and the matrix form of $\pi_{\mu\nu}$ may be more complicated than (2.181). The self-energy function $\pi_{\mu\nu}$ will be transverse but not necessarily perpendicular to $\mathbf{k}$. Then one must start from (2.180) and actually solve for the various components of $D_{\mu\nu}$.

In a homogeneous medium with dielectric constant ε, (2.169) is the equation obeyed by the vector potential. The double time derivative is multiplied by ε. Since this yields the ω^2

term in the Green's function, the appropriate Green's function in a medium with dielectric function ε is

$$D_{\mu\nu} = \frac{4\pi[\delta_{\mu\nu} - k_\mu k_\nu/k^2]}{\varepsilon\omega^2 - \omega_{\mathbf{k}}^2 + i\delta} \tag{2.185}$$

This equation for $D_{\mu\nu}$ is equated with Dyson's equation (2.184), for the dielectric function, which gives

$$\varepsilon = 1 - \frac{4\pi}{\omega^2}\pi^{(1)}(\mathbf{k}, \omega) \tag{2.186}$$

This result is not very useful for crystals, since there the form $\varepsilon_{\mu\nu}(\mathbf{k}, \omega)$ is a tensor, and is not the scalar function which we have given above. The scalar result is correct in crystals in the limit of $\mathbf{k} \to 0$:

$$\lim_{\mathbf{k}\to 0} \varepsilon_{\mu\nu}(\mathbf{k}, \omega) \to \delta_{\mu\nu}\left[1 - \frac{4\pi}{\omega^2}\pi^{(1)}(\mathbf{k}, \omega)\right] \tag{2.187}$$

It is interesting to note that in the limit of $\mathbf{k} \to 0$, this transverse dielectric function becomes exactly equal to the longitudinal one (2.171) at $\mathbf{k} \to 0$. This identity is not obvious yet and will be proved in Sec. 3.7.

PROBLEMS

1. Show explicitly that

$$\begin{aligned}\frac{1}{3!}\int_0^t dt_1 \int_0^t dt_2 \int_0^t dt_3 T[V(t_1)V(t_2)V(t_3)] \\ = \int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 V(t_1)V(t_2)V(t_3)\end{aligned} \tag{2.188}$$

2. For the phonon Green's function $D(\mathbf{q}, t - t')$, let $V(t)$ be the electron–phonon interaction and evaluate all the $n = 2$ diagrams. Which are connected, and which are disconnected? Draw the Feynman graphs for each term.

3. Let $V(t)$ be the electron–phonon interaction in the expansion for the electron Green's function $G(\mathbf{p}, t - t')$. What are the contributions from the different connected diagrams for $n = 4$ (two phonons). Just draw the graphs. Also draw all the graphs for the disconnected diagrams.

4. Let $V(t)$ be the electron–electron interaction in the expansion for the electron Green's function. Evaluate the term for $n = 1$, including all equations and Feynman graphs. Also draw the connected diagrams for $n = 2$.

5. Prove the Feynman result $e^{A+B} = e^A e^B e^{-(1/2)[A,B]}$, which is true only if $[A, B]$ commutes with both A and B. Hint: Recall that

$$e^{-it(H_0+V)} = e^{-itH_0} T \exp\left[-i\int_0^t dt_1 (e^{it_1 H_0} V e^{-it_1 H_0})\right] \tag{2.189}$$

Use the same method to prove that

$$e^{s(A+B)} = e^{sA} T_s \exp\left[\int_0^s ds_1 e^{-s_1 A} B e^{s_1 A}\right] \tag{2.190}$$

and evaluate the integral for $\mathbf{s} = 1$.

6. Evaluate the wave vector integrals at zero temperature for the exchange energy in (2.138).

7. Express $U_1(t) = e^{iH_0 t} e^{-iHt}$ and $U_2(t) = e^{iHt} e^{-iH_0 t}$ as time-ordered exponential integrals for both cases $t > 0$ and $t < 0$. You will need to use the operator T^{-1}, which arranges operators in their inverse time ordering.

8. Evaluate the first interaction term in Dyson's equation (2.160) and show that it does give (2.155) when $\Sigma(x_3, x_4)_{\mu\nu} = \delta^4(x_3 - x_4) V(x_3) \delta_{\mu\nu}$.

9. Show that Eq. (2.159) for G is identical to (2.118) using the result in (2.160). Self-energies obey the same relations as the Green's functions: i.e., $\Sigma_t = \Sigma^r + \Sigma^<$, etc.

10. Calculate the average value of the square of an ion displacement $\langle Q_j^2 \rangle$

$$\langle Q_j^2 \rangle = \frac{\hbar}{2\rho} \int \frac{d^d q}{(2\pi)^d} \frac{2N_q + 1}{\omega_q} \tag{2.191}$$

and show that it diverges in one and two dimensions due to long wavelength fluctuations.

Chapter 3

Nonzero Temperatures

3.1. INTRODUCTION

Experiments are done at nonzero temperatures. Since one goal of many-body theory is to explain experiments (another is to predict them), the theories should be done at nonzero temperatures too. It is often unnecessary if the temperature is small compared to other energies in the problem. But often temperature is important, and here it will be incorporated into Green's functions. The nonzero temperature formalism was originated by Matsubara (1955). It will actually be easier to use than the zero-temperature theory of Chapter 2, so that the Matsubara method will be used throughout the remainder of the book. The zero-temperature result is always easily obtained from the nonzero-temperature result by just setting $T = 0$.

At nonzero temperature, it is assumed there is something with a nonzero temperature. That is, our particle, whether electron, phonon, or spin, is interacting with a bath of other particles which have an average energy. The exact state of all these other particles is not known, since they are fluctuating between different configurations. All that is known is the temperature, which is related to the mean energy.

When defining the Green's function, one must average over all possible configurations of the system. A possible Green's function for the electron is

$$\frac{\mathrm{Tr}[e^{-\beta H} C_{\mathbf{p}\sigma}(t) C^{\dagger}_{\mathbf{p}\sigma}(t')]}{\mathrm{Tr}(e^{-\beta H})} \tag{3.1}$$

$$C_{\mathbf{p}\sigma}(t) = e^{itH} C_{\mathbf{p}\sigma} e^{-itH} \tag{3.2}$$

where the symbol "Tr" denotes trace and is the summation over some complete set of states:

$$\mathrm{Tr} \equiv \sum_n \langle n| \cdots |n\rangle \tag{3.3}$$

The definition (3.1) would be suitable for a Green's function and is $iG^{>}(\mathbf{p}; t, t')$. However, it has one drawback which makes its use unwieldy. Usually the Hamiltonian is written

$$H = H_0 + V \tag{3.4}$$

as a part H_0 that can be solved exactly and a part V which remains and becomes the perturbation. However, V now appears in two different places. First it is in $\exp(\pm iHt)$, which can be expanded in the usual S matrix. But it also occurs in the factor $\exp(-\beta H)$. There must

also be a perturbation expansion on the thermodynamic weighting factor. Of course, it is a nuisance to be doing two different expansions at once.

The Hamiltonian enters both terms as an exponential factor. The factor of $\beta = 1/k_B T$ can be considered to be a complex time. The Matsubara (1955) method does just the converse; it treats time as a complex temperature. The object is to treat t and β as the real and imaginary parts of a complex variable, which will require only one S-matrix expansion.

Another motivation for the Matsubara method is provided by examining the thermal occupation numbers for bosons $(e^{\beta\omega_q} - 1)^{-1}$ and fermions $(e^{\beta\xi_p} + 1)^{-1}$. Each of these can be expanded in a series ($\xi_p = \varepsilon_p - \mu$):

$$n_F(\xi_p) = \frac{1}{e^{\beta\xi_p} + 1} = \frac{1}{2} + \frac{1}{\beta}\sum_{n=-\infty}^{\infty} \frac{1}{(2n+1)i\pi/\beta - \xi_p} \tag{3.5}$$

$$n_B(\omega_q) = \frac{1}{e^{\beta\omega_q} - 1} = -\frac{1}{2} + \frac{1}{\beta}\sum_{n=-\infty}^{\infty} \frac{1}{2ni\pi/\beta - \omega_q} \tag{3.6}$$

These series can be derived from a theorem which states that any meromorphic function may be expanded as a summation over its poles and residues at those poles. The boson occupation factor $(e^{\beta\omega_q} - 1)^{-1}$ has poles at $\omega_q = 2\pi i n/\beta$ and the fermion factor $(e^{\beta\xi_p} + 1)^{-1}$ has poles at $\xi_p = (2n+1)i\pi/\beta$. It is convenient to define the frequencies at the pole

$$\begin{aligned} \omega_n &= (2n+1)\pi/\beta, \quad && \text{fermions} \\ \omega_n &= 2n\pi/\beta, && \text{bosons} \end{aligned} \tag{3.7}$$

where the fermions have poles at odd multiples of π/β, while bosons have poles at even multiples, including zero. Both summations above can be written as

$$\sum_n \frac{1}{i\omega_n - \omega_q} \quad \text{or} \quad \sum_n \frac{1}{i\omega_n - \xi_p} \tag{3.8}$$

where for fermions we sum over only odd integers and for bosons over even integers. The factor

$$\frac{1}{i\omega_n - \omega_q} \tag{3.9}$$

has the nature of a Green's function. Indeed, it is the noninteracting Green's function in the Matsubara method.

In the Matsubara method, time becomes a complex quantity which is usually called τ, where $\tau = it$. Green's functions are functions of τ with domain

$$-\beta \le \tau \le \beta \tag{3.10}$$

Fourier transform theory states that if a function $f(\tau)$ is defined over the range $-\beta \le \tau \le \beta$, then its Fourier expansion is

$$f(\tau) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}\left[a_n \cos\left(\frac{n\pi\tau}{\beta}\right) + b_n \sin\left(\frac{n\pi\tau}{\beta}\right)\right] \tag{3.11}$$

where

$$a_n = \frac{1}{\beta}\int_{-\beta}^{\beta} d\tau f(\tau)\cos\left(\frac{n\pi\tau}{\beta}\right) \tag{3.12}$$

$$b_n = \frac{1}{\beta}\int_{-\beta}^{\beta} d\tau f(\tau)\sin\left(\frac{n\pi\tau}{\beta}\right) \tag{3.13}$$

Another way to write the Fourier transform is to define

$$f(i\omega_n) = \tfrac{1}{2}\beta(a_n + ib_n) \tag{3.14}$$

and hence

$$f(\tau) = \frac{1}{\beta}\sum_{n=-\infty}^{\infty} e^{-in\pi\tau/\beta} f(i\omega_n) \tag{3.15}$$

$$f(i\omega_n) = \frac{1}{2}\int_{-\beta}^{\beta} d\tau f(\tau) e^{in\pi\tau/\beta} \tag{3.16}$$

There is still a further simplification which can be achieved. The boson Green's functions have the additional property that

$$\text{boson: } f(\tau) = f(\tau + \beta) \text{ when } -\beta < \tau < 0 \text{ (and } 0 < \tau + \beta < \beta) \tag{3.17}$$

Divide the integral (3.16) into its negative and positive regions:

$$f(i\omega_n) = \frac{1}{2}\left[\int_0^{\beta} d\tau f(\tau) e^{in\pi\tau/\beta} + \int_{-\beta}^{0} d\tau f(\tau) e^{in\pi\tau/\beta}\right] \tag{3.18}$$

and change variables in the second term from τ to $\tau + \beta$, which gives

$$f(i\omega_n) = \frac{1}{2}(1 + e^{in\pi})\int_0^{\beta} d\tau f(\tau) e^{in\pi/\beta} \tag{3.19}$$

This expression has the feature that $f(i\omega_n) = 0$ whenever n is an odd integer. For bosons,

$$\left.\begin{aligned} f(i\omega_n) &= \int_0^{\beta} d\tau e^{i\omega_n\tau} f(\tau) \\ f(\tau) &= \frac{1}{\beta}\sum_n e^{-i\omega_n\tau} f(i\omega_n) \\ \omega_n &= 2n\pi k_B T \end{aligned}\right\} \text{ bosons} \tag{3.20}$$

This result agrees with the previous observation (3.6) that boson frequencies contain only even integers.

Similarly, the fermion Green's functions will have the property that

$$\text{fermions: } f(\tau) = -f(\tau + \beta) \text{ when } -\beta < \tau < 0 \tag{3.21}$$

The same manipulations on the integral in (3.16) give

$$f(i\omega_n) = \frac{1}{2}(1 - e^{in\pi})\int_0^{\beta} d\tau f(\tau) e^{in\pi/\beta} \tag{3.22}$$

In this case $f(i\omega_n) = 0$ if n is even, while for n an odd integer,

$$\left.\begin{aligned} f(i\omega_n) &= \int_0^\beta d\tau e^{i\omega_n\tau} f(\tau) \\ f(\tau) &= \frac{1}{\beta}\sum_n e^{-i\omega_n\tau} f(i\omega_n) \\ \omega_n &= (2n+1)\pi k_B T \end{aligned}\right\} \text{fermions} \tag{3.23}$$

These equations are identical in form to those in (3.20). The only difference is whether the frequency ω_n has even or odd integers. This pair of equations will be used often to define the Fourier transforms of Green's functions.

At this point the reader is probably overwhelmed with complex frequencies and complex times. One might argue that this suffering is necessary when including nonzero temperatures. Actually, the forecast is not as gloomy as that. You will find the Matsubara method is very easy to use. It is particularly good for evaluating high-order diagrams with many internal lines. In fact, it is a remarkably easy method to use in practice. Except for the notion of complex times and frequencies, it is not that much different from the zero-temperature Green's functions which we defined in Chapter 2.

A second great merit of the Matsubara method is that it leads us directly to physical results. In Secs. 3.7 and 3.8, some Kubo formulas are derived for the exact definitions of physical quantities such as the electrical conductivity, thermal conductivity, magnetic susceptibility, etc. In Sec. 3.3, it will be shown that these correlation functions are just retarded Green's functions. Finally, it will be shown that the Matsubara Green's functions lead directly to the retarded functions. Our Matsubara functions will be functions of the complex frequencies $i\omega_n$, such as $f(i\omega_n)$. The equivalent retarded function is obtained by replacing $i\omega_n$ by $\omega + i\delta$, where δ is infinitesimal and $i = \sqrt{-1}$. This step is called an analytical continuation. In practice, one just takes the formula one has derived for $f(i\omega_n)$, erases $i\omega_n$ everywhere, and replaces it by $\omega + i\delta$. This simple procedure yields the retarded function, which is needed for the physically measurable quantities. The Matsubara technique is a direct method of calculating the quantities which can be compared with experiment.

3.2. MATSUBARA GREEN'S FUNCTIONS

The electron Green's function is defined as

$$\mathcal{G}(\mathbf{p}, \tau - \tau') = -\langle T_\tau C_{\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle \tag{3.24}$$

$$\begin{aligned} \mathcal{G}(\mathbf{p}, \tau - \tau') = -\mathrm{Tr}[&e^{-\beta(H-\mu N-\Omega)} T_\tau e^{\tau(H-\mu N)} C_{\mathbf{p}\sigma} e^{-(\tau-\tau')(H-\mu N)} \\ &\times C^\dagger_{\mathbf{p}\sigma} e^{-\tau'(H-\mu N)}] \end{aligned} \tag{3.25}$$

$$e^{-\beta\Omega} = \mathrm{Tr}(e^{-\beta(H-\mu N)}) \tag{3.26}$$

These definitions have several features and conventions which need to be explained. First, the bracket $\langle\cdots\rangle$ in (3.24) has the definition implied by the equivalent equation (3.25). The bracket $\langle O\rangle$ on an operator O means to take the thermodynamic average, which is the trace over the complete set of states. Second, the Hamiltonian is now replaced by $H - \mu N$, where μ is the chemical potential and N is the particle number operator. A grand canonical ensemble is used, where the number of particles is variable. This definition of the Green's function applies

to a many-particle system. It can also be used very successfully for one particle in an empty band. In the latter case, the analytical continuation is taken as $i\omega_n \to E + \mu + i\delta$, and the chemical potential will vanish from all expressions. One is not bothered by the fact that $\beta\mu \ll 0$ in one-particle systems at nonzero temperatures.

In a many-electron system, the chemical potential is retained in the formalism. The analytical continuation is $i\omega_n \to E + i\delta$ and energy is measured from the chemical potential (Fermi energy). The factor T_τ is a τ-ordering operator, which arranges operators with earliest τ (closest to $-\beta$) to the right. It serves the same function as the time-ordering operator in the zero-temperature Green's functions. The subscript τ is affixed to T to distinguish this operator from the temperature. The thermodynamic potential Ω in $\exp(-\beta\Omega)$ is the usual normalization factor for a thermodynamic average. The script symbol $\mathcal{G}$ has been used for these Matsubara functions. This script will always be written to alert the reader that these are Green's functions of complex time and complex frequency.

In (3.24) the Green's function on the left has been written as a function of the difference $\tau - \tau'$, although the right-hand side is not obviously a function of only the difference. Now prove this to be the case. First, write the Green's function for the separate cases for $\tau > \tau'$ and $\tau < \tau'$:

$$K \equiv H - \mu N \tag{3.27}$$

$$\begin{aligned}\mathcal{G}(\mathbf{p}, \tau - \tau') = &-\Theta(\tau - \tau')\mathrm{Tr}[e^{-\beta(K-\Omega)}e^{\tau K}C_{\mathbf{p}\sigma}e^{-(\tau-\tau')K}C^\dagger_{\mathbf{p}\sigma}e^{-\tau' K}] \\ &+ \Theta(\tau' - \tau)\mathrm{Tr}[e^{-\beta(K-\Omega)}e^{\tau' K}C^\dagger_{\mathbf{p}\sigma}e^{(\tau-\tau')K}C_{\mathbf{p}\sigma}e^{-\tau K}]\end{aligned} \tag{3.28}$$

The sign change in the second term appears whenever two fermion operators are interchanged. Next, use the theorem that the trace is unchanged by a cyclic variation of the operators

$$\mathrm{Tr}(ABC\cdots YZ) = \mathrm{Tr}(BC\cdots XYZA) \tag{3.29}$$

to move the factors of $\exp(-\tau^{(\prime)}K)$ to the left. So Eq. (3.28) can be rewritten as

$$\begin{aligned}\mathcal{G}(\mathbf{p}, \tau - \tau') = &-\Theta(\tau - \tau')\mathrm{Tr}([e^{-\tau' K}e^{-\beta(K-\Omega)}e^{\tau K}C_{\mathbf{p}\sigma}e^{-(\tau-\tau')K}C^\dagger_{\mathbf{p}\sigma}] \\ &+ \Theta(\tau' - \tau)\mathrm{Tr}[e^{-\tau K}e^{-\beta(K-\Omega)}e^{\tau' K}C^\dagger_{\mathbf{p}\sigma}e^{(\tau-\tau')K}C_{\mathbf{p}\sigma}]\end{aligned} \tag{3.30}$$

Next commute the exponential operators,

$$e^{-\tau' K}e^{-\beta(K-\Omega)} = e^{-\beta(K-\Omega)}e^{-\tau' K} \tag{3.31}$$

since they both contain the same operator K [the thermodynamic potential Ω is not an operator but is a scalar function of β and μ, as defined in (3.26)]:

$$\begin{aligned}\mathcal{G}(\mathbf{p}, \tau - \tau') = &-\Theta(\tau - \tau')\mathrm{Tr}[e^{-\beta(K-\Omega)}e^{(\tau-\tau')K}C_{\mathbf{p}\sigma}e^{-(\tau-\tau')K}C^\dagger_{\mathbf{p}\sigma}] \\ &+ \Theta(\tau' - \tau)\mathrm{Tr}[e^{-\beta(K-\Omega)}e^{-(\tau-\tau')K}C^\dagger_{\mathbf{p}\sigma}e^{(\tau-\tau')K}C_{\mathbf{p}\sigma}]\end{aligned} \tag{3.32}$$

The right-hand side of this equation is now a function only of the combination $\tau - \tau'$. The Green's function can always be written as a function of this difference. It enables us to drop one of the time variables since it is unnecessary. An equivalent definition of the Green's function is

$$\mathcal{G}(\mathbf{p}, \tau) = -\langle T_\tau C_{\mathbf{p}\sigma}(\tau)C^\dagger_{\mathbf{p}\sigma}(0)\rangle \tag{3.33}$$

$$= -\mathrm{Tr}[e^{-\beta(K-\Omega)}T_\tau(e^{\tau K}C_{\mathbf{p}\sigma}e^{-\tau K}C^\dagger_{\mathbf{p}\sigma})] \tag{3.34}$$

Next the Green's function is examined for $\tau < 0$ to verify that it does have the property asserted in (3.21):

$$\tau < 0: \quad \mathcal{G}(\mathbf{p}, \tau) = \mathrm{Tr}(e^{-\beta(K-\Omega)} C^\dagger_{\mathbf{p}\sigma} e^{\tau K} C_{\mathbf{p}\sigma} e^{-\tau K}) \tag{3.35}$$

By using the cyclic property of the trace several times, the above equation can be rearranged into

$$\tau < 0: \quad \mathcal{G}(\mathbf{p}, \tau) = \mathrm{Tr}(e^{\beta\Omega} e^{\tau K} C_{\mathbf{p}\sigma} e^{-(\tau+\beta)K} C^\dagger_{\mathbf{p}\sigma}) \tag{3.36}$$

The factor of $\exp(\beta\Omega)$ is not cycled since it is not an operator. These terms can be regrouped by adding $\exp(\pm\beta K)$ to the first terms to give

$$\tau < 0: \quad \mathcal{G}(\mathbf{p}, \tau) = \mathrm{Tr}(e^{-\beta(K-\Omega)} e^{(\tau+\beta)K} C_{\mathbf{p}\sigma} e^{-(\tau+\beta)K} C^\dagger_{\mathbf{p}\sigma}) \tag{3.37}$$

The term on the right is $-\mathcal{G}(\mathbf{p}, \tau + \beta)$ when $0 < \tau + \beta < \beta$. These steps show that

$$-\beta < \tau < 0: \quad \mathcal{G}(\mathbf{p}, \tau) = -\mathcal{G}(\mathbf{p}, \tau + \beta) \tag{3.38}$$

as was asserted earlier in (3.21). The above identity proves that the Green's function can be expanded in a Fourier series of the type in (3.23):

$$\mathcal{G}(\mathbf{p}, i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} \mathcal{G}(\mathbf{p}, \tau) \tag{3.39}$$

$$\mathcal{G}(\mathbf{p}, \tau) = \frac{1}{\beta} \sum_n e^{-i\omega_n \tau} \mathcal{G}(\mathbf{p}, i\omega_n) \tag{3.40}$$

Equation (3.39) serves as the definition of $\mathcal{G}(\mathbf{p}, i\omega_n)$, where $i\omega_n$ is always an odd multiple of π/β for fermions.

The noninteracting Green's function, or free-particle Green's function, is obtained from (3.34) by using for the Hamiltonian

$$H = H_0 = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} \tag{3.41}$$

$$K = K_0 = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} \tag{3.42}$$

$$\xi_{\mathbf{p}} = \varepsilon_{\mathbf{p}} - \mu \tag{3.43}$$

The τ evolution of the operators is just

$$C_{\mathbf{p}\sigma}(\tau) = e^{\tau K_0} C_{\mathbf{p}\sigma} e^{-\tau K_0} = e^{-\xi_{\mathbf{p}}\tau} C_{\mathbf{p}\sigma} \tag{3.44}$$

$$C^\dagger_{\mathbf{p}\sigma}(\tau) = e^{\tau K_0} C^\dagger_{\mathbf{p}\sigma} e^{-\tau K_0} = e^{\xi_{\mathbf{p}}\tau} C^\dagger_{\mathbf{p}\sigma} \tag{3.45}$$

which is easily derived from the Baker–Hausdorff theorem:

$$e^A C e^{-A} = C + [A, C] + \frac{1}{2!}[A, [A, C]] + \frac{1}{3!}[A, [A, [A, C]]] + \cdots \tag{3.46}$$

The τ dependence of the Green's function is

$$\mathcal{G}^{(0)}(\mathbf{p}, \tau) = -\Theta(\tau) e^{-\xi_{\mathbf{p}}\tau} \langle C_{\mathbf{p}\sigma} C^\dagger_{\mathbf{p}\sigma} \rangle + \Theta(-\tau) e^{-\xi_{\mathbf{p}}\tau} \langle C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} \rangle \tag{3.47}$$

$$= -e^{-\xi_{\mathbf{p}}\tau} \{\Theta(\tau)[1 - n_F(\xi_{\mathbf{p}})] - \Theta(-\tau) n_F(\xi_{\mathbf{p}})\} \tag{3.48}$$

$$= -e^{-\xi_{\mathbf{p}}\tau} [\Theta(\tau) - n_F(\xi_{\mathbf{p}})] \tag{3.49}$$

where $n_F(\xi_{\mathbf{p}})$ is the expectation of the number operator: $n_F(\xi_{\mathbf{p}}) = \langle C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma}\rangle$, which from elementary statistical mechanics has the form

$$n_F(\xi_{\mathbf{p}}) = \frac{1}{e^{\beta\xi_{\mathbf{p}}} + 1} \tag{3.50}$$

It is also easy to obtain the Green's function of frequency:

$$\mathcal{G}^{(0)}(\mathbf{p}, i\omega_n) = \int_0^\beta d\tau e^{i\omega_n\tau}\mathcal{G}^{(0)}(\mathbf{p}, \tau) = -(1 - n_F)\int_0^\beta d\tau e^{\tau(i\omega_n - \xi_{\mathbf{p}})}$$

$$\mathcal{G}^{(0)}(\mathbf{p}, i\omega_n) = -\frac{(1 - n_F)(e^{\beta(i\omega_n - \xi_{\mathbf{p}})} - 1)}{i\omega_n - \xi_{\mathbf{p}}} \tag{3.51}$$

The second term in the numerator may be simplified by remembering that

$$i\beta\omega_n = i(2n + 1)\pi \tag{3.52}$$

$$e^{i\beta\omega_n} = -1 \tag{3.53}$$

which gives

$$\mathcal{G}^{(0)}(\mathbf{p}, i\omega_n) = \frac{(1 - n_F)(e^{-\beta\xi_{\mathbf{p}}} + 1)}{i\omega_n - \xi_{\mathbf{p}}} \tag{3.54}$$

$$\mathcal{G}^{(0)}(\mathbf{p}, i\omega_n) = \frac{1}{i\omega_n - \xi_{\mathbf{p}}} \tag{3.55}$$

since it is easy to show from (3.50) that $[1 - n_F] = 1/(e^{-\beta\xi_{\mathbf{p}}} + 1)$. Equation (3.55) is the noninteracting Green's function for electrons.

The result for $\mathcal{G}^{(0)}$ does have the form suggested in Sec. 3.1. Temperature information is still contained in this expression but now only in the frequency $(2n + 1)\pi/\beta$. Later it is shown how the occupation factors reenter the expressions when the diagrams and correlation functions are actually evaluated.

The phonon and photon Green's functions are defined in the same fashion. They are obviously similar to each other, so only the derivation of the phonon Green's function is presented. The photon results will be stated at the end. For phonons, in the interval $-\beta \le \tau \le \beta$ the Green's function is

$$\mathcal{D}(\mathbf{q}, \tau - \tau') = -\langle T_\tau A(\mathbf{q}, \tau)A(-\mathbf{q}, \tau')\rangle \tag{3.56}$$

$$A(\mathbf{q}, \tau) = e^{\tau H}(a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}})e^{-\tau H} \tag{3.57}$$

Phonons have no chemical potential, since one can make an arbitrary number of them, and the tau dependence is governed just by the Hamiltonian. Again one can show that the right-hand side of (3.56) is only a function of $\tau - \tau'$. It is not necessary to keep two τ variables, and one can instead define

$$\mathcal{D}(\mathbf{q}, \tau) = -\langle T_\tau A(\mathbf{q}, \tau)A(-\mathbf{q}, 0)\rangle \tag{3.58}$$

Next, examine the behavior for negative tau:

$$\tau < 0\colon\ \mathcal{D}(\mathbf{q}, \tau) = -\langle A(-\mathbf{q}, 0)A(\mathbf{q}, \tau)\rangle \tag{3.59}$$

$$= -\mathrm{Tr}[e^{-\beta(H-\Omega)}A(-\mathbf{q})e^{\tau H}A(\mathbf{q})e^{-\tau H}] \tag{3.60}$$

Using cyclic permutations of the trace turns the trace into

$$\tau < 0: \quad D(\mathbf{q}, \tau) = -\mathrm{Tr}[e^{\beta\Omega} e^{\tau H} A(\mathbf{q}) e^{-(\tau+\beta)H} A(-\mathbf{q})] \tag{3.61}$$

$$\tau < 0: \quad D(\mathbf{q}, \tau) = -\mathrm{Tr}[e^{-\beta(H-\Omega)} e^{(\tau+\beta)H} A(\mathbf{q}) e^{-(\tau+\beta)H} A(-\mathbf{q})] \tag{3.62}$$

which proves

$$-\beta < \tau < 0: \quad D(\mathbf{q}, \tau) = D(\mathbf{q}, \tau + \beta) \tag{3.63}$$

The right-hand side of the equation is the Green's function when $0 < \tau + \beta < \beta$. This identity satisfies the general conditions asserted in (3.17) for boson correlation functions. The Fourier transform has the following form in (3.20):

$$\mathcal{D}(\mathbf{q}, i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} \mathcal{D}(\mathbf{q}, \tau) \tag{3.64}$$

$$\mathcal{D}(\mathbf{q}, \tau) = \frac{1}{\beta} \sum_n e^{-i\omega_n \tau} \mathcal{D}(\mathbf{q}, i\omega_n) \tag{3.65}$$

$$\omega_n = 2n\pi k_B T \tag{3.66}$$

Equation (3.65) provides the definition of the frequency-dependent Green's function.

The difference between (3.17) and (3.21) is just a sign change. The fermion functions have a sign change because their operators obey anti-commutation relations, while the bosons have no sign change because their operators obey commutation relations. Of course, this sign change is a result of the fundamental difference between bosons and fermions. The sign change is also responsible for the sign change between ± 1 in their two forms of thermal distributions: $(e^{\beta\xi_{\mathbf{p}}} + 1)^{-1}$ vs. $(e^{\beta\omega_{\mathbf{q}}} - 1)^{-1}$. This sign change is very important, and one has to keep track of it carefully in fermion problems with many operators.

For phonons, the noninteracting or free-phonon Green's function is obtained by taking $H = H_0 = \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}}$, which for the τ variation of the operators yields

$$a_{\mathbf{q}}(\tau) = e^{\tau H_0} a_{\mathbf{q}} e^{-\tau H_0} = e^{-\tau\omega_{\mathbf{q}}} a_{\mathbf{q}} \tag{3.67}$$

$$a_{\mathbf{q}}^\dagger(\tau) = e^{\tau H_0} a_{\mathbf{q}}^\dagger e^{-\tau H_0} = e^{\tau\omega_{\mathbf{q}}} a_{\mathbf{q}}^\dagger \tag{3.68}$$

Always remember that $[a_{\mathbf{q}}(\tau)]^\dagger \neq a_{\mathbf{q}}^\dagger(\tau)$. The noninteracting Green's function is

$$\begin{aligned} \mathcal{D}^{(0)}(\mathbf{q}, \tau) = &-\Theta(\tau)\langle (a_{\mathbf{q}} e^{-\tau\omega_{\mathbf{q}}} + a_{-\mathbf{q}}^\dagger e^{\tau\omega_{\mathbf{q}}})(a_{-\mathbf{q}} + a_{\mathbf{q}}^\dagger)\rangle \\ &- \Theta(-\tau)\langle (a_{-\mathbf{q}} + a_{\mathbf{q}}^\dagger)(a_{\mathbf{q}} e^{-\tau\omega_{\mathbf{q}}} + a_{-\mathbf{q}}^\dagger e^{\tau\omega_{\mathbf{q}}})\rangle \end{aligned} \tag{3.69}$$

Capital letters are used to signify the thermal expectation value of boson number operators:

$$N_{\mathbf{q}} = \langle a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \rangle = n_B(\omega_{\mathbf{q}}) = \frac{1}{e^{\beta\omega_{\mathbf{q}}} - 1} \tag{3.70}$$

$$N_{\mathbf{q}} + 1 = \langle a_{\mathbf{q}} a_{\mathbf{q}}^\dagger \rangle \tag{3.71}$$

Averages such as $\langle a_{\mathbf{q}} a_{\mathbf{q}} \rangle$ and $\langle a_{\mathbf{q}}^\dagger a_{\mathbf{q}}^\dagger \rangle$ yield zero since they vanish for each term in the trace.

The Green's function of τ can be written as

$$\begin{aligned} \mathcal{D}^{(0)}(\mathbf{q}, \tau) = &-\Theta(\tau)[(N_{\mathbf{q}} + 1)e^{-\tau\omega_{\mathbf{q}}} + N_{\mathbf{q}} e^{\tau\omega_{\mathbf{q}}}] \\ &- \Theta(-\tau)[N_{\mathbf{q}} e^{-\tau\omega_{\mathbf{q}}} + (N_{\mathbf{q}} + 1)e^{-\tau\omega_{\mathbf{q}}}] \end{aligned} \tag{3.72}$$

The Green's function of frequency is

$$\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n) = \int_0^\beta d\tau e^{i\omega_n\tau}\mathcal{D}^{(0)}(\mathbf{q}, \tau)$$
$$= -\left[(N_\mathbf{q}+1)\frac{(e^{\beta(i\omega_n-\omega_\mathbf{q})}-1)}{i\omega_n-\omega_\mathbf{q}} + N_\mathbf{q}\frac{(e^{\beta(i\omega_n+\omega_\mathbf{q})}-1)}{i\omega_n+\omega_\mathbf{q}}\right] \tag{3.73}$$

The terms in the numerators may be simplified by noting that for bosons $\exp(i\omega_n\beta) = 1$ so that the Green's function is

$$\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n) = -\left[(N_\mathbf{q}+1)\frac{(e^{-\beta\omega_\mathbf{q}}-1)}{i\omega_n-\omega_\mathbf{q}} + N_\mathbf{q}\frac{(e^{\beta\omega q}-1)}{i\omega_n+\omega_\mathbf{q}}\right] \tag{3.74}$$

Using Eq. (3.70) shows that the first numerator equals -1, and the second is $+1$:

$$\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n) = -\left(-\frac{1}{i\omega_n-\omega_\mathbf{q}} + \frac{1}{i\omega_n+\omega_\mathbf{q}}\right) \tag{3.75}$$

$$\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n) = \frac{2\omega_\mathbf{q}}{(i\omega_n)^2-\omega_\mathbf{q}^2} = -\frac{2\omega_\mathbf{q}}{\omega_n^2+\omega_\mathbf{q}^2} \tag{3.76}$$

This Green's function has a simple form. It is almost identical to the zero-temperature case (2.72), and the only difference is the use of complex frequencies instead of real ones. Note that $\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n)$ is always negative.

The photon Green's function is also identical to its zero-temperature result, except for complex frequencies. The fundamental definition is

$$\mathcal{D}_{\mu\nu}(\mathbf{k}, \tau) = -\sum_\lambda \langle T_\tau A_\mu(\mathbf{k}, \lambda, \tau)A_\nu(-\mathbf{k}, \lambda, 0)\rangle \tag{3.77}$$

$$A_\mu(\mathbf{k}, \lambda, 0) = \xi_\mu(\mathbf{k}, \lambda)\left(\frac{2\pi}{\omega_\mathbf{k}}\right)(a_{\mathbf{k}\lambda} + a^\dagger_{\mathbf{k}\lambda}) \tag{3.78}$$

where the operator A_μ is the usual vector potential operator in (2.163). The free-photon Green's function is

$$\mathcal{D}^{(0)}_{\mu\nu}(\mathbf{k}, i\omega_n) = -\frac{4\pi(\delta_{\mu\nu} - k_\mu k_\nu/k^2)}{\omega_n^2+\omega_\mathbf{k}^2} \tag{3.79}$$

which should be compared with (2.178).

This section is ended with a comment on notation. The following three forms for the Green's function are equivalent and will be used interchangeably:

$$\mathcal{G}(\mathbf{p}, ip_n) = \mathcal{G}(\mathbf{p}, ip) = \mathcal{G}(p) \tag{3.80}$$

$$\mathcal{D}(\mathbf{q}, i\omega_n) = \mathcal{D}(\mathbf{q}, i\omega) = \mathcal{D}(q) \tag{3.81}$$

The form on the left has been used so far. In the second form, ip has been used instead of ip_n. They mean the same thing, since the i in ip is enough information to alert the reader that complex frequencies are being employed, which are always discrete. Hence the n subscript is redundant. In the last form, a four-vector notation $p = (\mathbf{p}, ip)$ is used, and the script form of $\mathcal{G}$ is sufficient to alert the reader to the use of Matsubara Green's functions.

3.3. RETARDED AND ADVANCED GREEN'S FUNCTIONS

The retarded and advanced Green's functions were introduced in Sec. 2.9. They play an important role in the nonzero-temperature theory. Their properties are discussed in this section. Their importance comes from the fact that all measurable quantities, such as conductivities or susceptibilities, are actually retarded correlation functions. The goal of many calculations is to calculate a retarded function. There are several different ways to obtain them. One is to use a real time theory even at nonzero temperatures. This method is obvious to the beginner, but is the hardest way. The second method, which is used most often, is to first calculate the equivalent Matsubara function of imaginary frequency. It is shown below that the retarded function is obtained from the Matsubara function by simply changing $i\omega_n$ to $\omega + i\delta$ where δ is infinitesimal. The Matsubara function is the easiest one to calculate because its S-matrix expansion is simple. The retarded function is most easily found from the Matsubara function.

The retarded Green's functions may be defined for both zero and nonzero temperatures. The retarded Green's function for an electron in state $\mathbf{p}$ is

$$\begin{aligned} G_{\text{ret}}(\mathbf{p}, t-t') &= -i\Theta(t-t')\langle[C_{\mathbf{p}\sigma}(t)C^\dagger_{\mathbf{p}\sigma}(t') + C^\dagger_{\mathbf{p}\sigma}(t')C_{\mathbf{p}\sigma}(t)]\rangle \\ &= -i\Theta(t-t')\text{Tr}\{e^{-\beta(K-\Omega)}[C_{\mathbf{p}\sigma}(t)C^\dagger_{\mathbf{p}\sigma}(t') + C^\dagger_{\mathbf{p}\sigma}(t')C_{\mathbf{p}\sigma}(t)]\} \qquad (3.82) \end{aligned}$$

$$K = H - \mu N, \;\; C_{\mathbf{p}\sigma}(t) = e^{iKt}C_{\mathbf{p}\sigma}e^{-itK} \qquad (3.83)$$

The brackets $\langle\cdots\rangle$ indicate thermodynamic average, as is explicitly shown on the second line. The square brackets mean nothing in particular; they are used to group symbols together. The retarded Green's function depends on real time, not tau. The tip-off for this is the i factor in front which belongs with all real-time Green's functions. The Green's function operates only for $t > t'$, which makes it causal. One starts a signal at one time t' and measures it later at t. Of course, actual systems are causal, which is why these Green's functions are the ones of physical interest. The argument of the Green's function is an anticommutator at different times. In the limit that the times become equal, the anticommutator becomes unity,

$$1 = \lim_{t\to t'}\{C_{\mathbf{p}\sigma}(t)C^\dagger_{\mathbf{p}\sigma}(t') + C^\dagger_{\mathbf{p}\sigma}(t')C_{\mathbf{p}\sigma}(t)\} \qquad (3.84)$$

since it just becomes the usual fermion anticommutator. The plus sign in the middle of the two terms is an important feature for retarded Green's functions of fermion operators. The left-hand side of (3.82) indicates that the retarded function depends only upon the difference $(t - t')$. This feature can be shown by manipulations on the trace similar to those in the last section.

For phonons, the retarded Green's function is

$$D_{\text{ret}}(\mathbf{q}, t-t') = -i\Theta(t-t')\langle A(\mathbf{q}, t)A(-\mathbf{q}, t') - A(-\mathbf{q}, t')A(\mathbf{q}, t)\rangle \qquad (3.85)$$

It is very similar to (3.82) in that it is for real time, is also thermodynamically averaged, and is defined only for $t > t'$. However, the sign in the middle is now minus, which corresponds to the fact that bosons obey commutation relations. For both electron and phonon retarded functions, the right-hand side can be shown to be a function only of $t - t'$, as is indicated in the argument of the Green's function on the left-hand side of the definition.

Retarded Green's functions are needed for many types of operators. These operators will usually be products of electron or boson operators. For example, let us define the operators

$$U = \sum_{ij} M_{ij} C_i^\dagger C_j \tag{3.86}$$

$$V = \sum_{ijk} M_{ijk} C_i^\dagger C_j C_k \tag{3.87}$$

The operator U is bilinear in the operators C_i, where M_{ij} is just a matrix element. The operator U is regarded as having boson properties, regardless of whether both Cs are fermion or both boson operators; the case is excluded where one C is boson and one is fermion. U is boson because it acts as a composite particle. This bilinear form will be encountered quite often, since it is characteristic of some important operators such as the current and density operators. The retarded Green's function for the operator U is denoted $\bar{U}$ and is defined as

$$\bar{U}_{\text{ret}}(t - t') = -i\Theta(t - t')\langle[U(t)U^\dagger(t') - U^\dagger(t')U(t)]\rangle \tag{3.88}$$

This definition is similar to (3.85), with the important feature that it has the minus sign in the center of the bracket, which is the case for all boson operators; that is, for any operator which is a product of bosons or an even number of fermions.

However, an operator such as V above is considered fermion if it is a product of an odd number of fermions. Its retarded function is

$$\bar{V}_{\text{ret}}(t - t') = -i\Theta(t - t')\langle[V(t)V^\dagger(t') + V^\dagger(t')V(t)]\rangle \tag{3.89}$$

which now has the plus sign in the bracket.

All these retarded functions have Fourier transforms defined by the usual convention:

$$G_{\text{ret}}(\mathbf{p}, E) = \int_{-\infty}^{\infty} dt e^{iE(t-t')} G_{\text{ret}}(\mathbf{p}, t - t') \tag{3.90}$$

$$D_{\text{ret}}(\mathbf{q}, \omega) = \int_{-\infty}^{\infty} dt e^{i\omega(t-t')} D_{\text{ret}}(\mathbf{q}, t - t') \tag{3.91}$$

$$\bar{U}_{\text{ret}}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} \bar{U}_{\text{ret}}(t) \tag{3.92}$$

The advanced Green's function for each of these is defined by

$$G_{\text{adv}}(\mathbf{p}, t - t') = i\Theta(t' - t)\langle[C_{\mathbf{p}}(t)C_{\mathbf{p}}^\dagger(t') + C_{\mathbf{p}}^\dagger(t')C_{\mathbf{p}}(t)]\rangle \tag{3.93}$$

$$\begin{aligned} D_{\text{adv}}(\mathbf{q}, t - t') = i\Theta(t' - t)\langle &A(\mathbf{q}, t)A(-\mathbf{q}, t') \\ &- A(-\mathbf{q}, t')A(\mathbf{q}, t)\rangle \end{aligned} \tag{3.94}$$

$$\bar{U}_{\text{adv}}(t - t') = i\Theta(t' - t)\langle[U(t)U^\dagger(t') - U^\dagger(t')U(t)]\rangle \tag{3.95}$$

The only two differences are the sign change in front and the fact that the time domain is now $t' > t$, which is just opposite of that for retarded functions. Their Fourier transforms with respect to frequency are defined in the usual way as in (3.91).

The advanced functions of energy turn out to be the complex conjugate of the corresponding retarded function. To prove this fact, first start with the advanced function, then take its Hermitian conjugate, and finally invert its time variables. The result is the retarded function

$$\bar{U}_{\text{adv}}(t'-t) = i\Theta(t-t')\langle[U(t')U^\dagger(t) - U^\dagger(t)U(t')]\rangle \tag{3.96}$$

$$\bar{U}_{\text{adv}}(t'-t)^\dagger = \bar{U}_{\text{ret}}(t-t') \tag{3.97}$$

Now take the Fourier transform of both sides,

$$\bar{U}_{\text{ret}}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega(t-t')}\bar{U}_{\text{adv}}(t'-t)^\dagger = \int_{-\infty}^{\infty} dt_1 e^{-i\omega t_1}\bar{U}_{\text{adv}}(t_1)^\dagger$$

where the last step is obtained by changing variables $t_1 = t' - t$. These steps prove that

$$\bar{U}_{\text{ret}}(\omega) = \bar{U}_{\text{adv}}(\omega)^* \tag{3.98}$$

This result can be generalized to any of the retarded and advanced Green's functions. It is sufficient to find the retarded function, since a simple complex conjugation derives the advanced one.

A particular representation for these Green's functions is introduced. This representation is a formal one, which is not generally useful for calculating physical quantities and determining numbers. However, it is useful for proving theorems and in particular for relating one Green's function to another. This representation uses the complete set of states $|m\rangle$ which are the exact eigenstates of $K = H - \mu N$. Usually the exact eigenvalues and eigenstates of K are not known. However, in principle they exist, which is sufficient for proving theorems. The eigenvalues of K are denoted E_m

$$K|m\rangle = E_m|m\rangle \tag{3.99}$$

This complete set of states will be used in the thermodynamic average: the symbol Tr denotes trace, and the set $|n\rangle$ are used for this summation:

$$\bar{U}_{\text{ret}}(t-t') = -i\Theta(t-t')e^{\beta\Omega}\sum_n \langle n|e^{-\beta K}[U(t) \wedge U^\dagger(t') - U^\dagger(t') \wedge U(t)]|n\rangle$$

The above equation has two points marked by an insertion sign "$\wedge$". In both places, a complete set of states is inserted which is unity,

$$1 = \sum_m |m\rangle\langle m| \tag{3.100}$$

which gives

$$\bar{U}_{\text{ret}}(t-t') = -i\Theta(t-t')e^{\beta\Omega}\sum_{m,n} e^{-\beta E_n}[\langle n|U(t)|m\rangle\langle m|U^\dagger(t')|n\rangle - \langle n|U^\dagger(t')|m\rangle\langle m|U(t)|n\rangle] \tag{3.101}$$

The matrix elements are easy to evaluate

$$\langle n|U(t)|m\rangle = \langle n|e^{itK}Ue^{-itK}|m\rangle = \langle n|U|m\rangle e^{it(E_n - E_m)} \tag{3.102}$$

which gives for the retarded function

$$\bar{U}_{\text{ret}}(t-t') = -i\Theta(t-t')e^{\beta\Omega}\sum_{m,n} e^{-\beta E_n}[e^{i(t-t')(E_n-E_m)}|\langle n|U(t)|m\rangle|^2 - e^{-i(t-t')(E_n-E_m)}|\langle m|U(t)|n\rangle|^2] \tag{3.103}$$

In the second term exchange the dummy summation variables n and m so that the matrix elements $\langle n|U|m\rangle$ are the same in each term:

$$\bar{U}_{\text{ret}}(t-t') = -i\Theta(t-t')e^{\beta\Omega}\sum_{m,n}|\langle n|U|m\rangle|^2 e^{i(t-t')(E_n-E_m)} \times [e^{-\beta E_n} - e^{-\beta E_m}]$$

This formula is the result for the retarded Green's function of time. The Fourier transform is taken to give the frequency function,

$$\bar{U}_{\text{ret}}(\omega) = -i\int_0^\infty e^{it(\omega+i\delta)}\,dt e^{\beta\Omega}\sum_{m,n}|\langle n|U|m\rangle|^2 e^{it(E_n-E_m)}[e^{-\beta E_n} - e^{-\beta E_m}]$$
$$= e^{\beta\Omega}\sum_{m,n}|\langle n|U|m\rangle|^2\frac{e^{-\beta E_n} - e^{-\beta E_m}}{\omega + E_n - E_m + i\delta} \tag{3.104}$$

where $i\delta$ is added to the frequency to ensure convergence at large times.

The equivalent Matsubara function for the operator U is defined by a script symbol:

$$\mathcal{U}(\tau) = -\langle T_\tau U(\tau)U^\dagger(0)\rangle \tag{3.105}$$

$$\mathcal{U}(i\omega_n) = \int_0^\beta d\tau e^{i\omega_n\tau}\mathcal{U}(\tau) \tag{3.106}$$

The representation $|n\rangle$ is also applied to this expression

$$\tau > 0: \quad \mathcal{U}(\tau) = -e^{\beta\Omega}\sum_{n,m}\langle n|e^{-\beta K}U(\tau)|m\rangle\langle m|U^\dagger(0)|n\rangle \tag{3.107}$$

$$\mathcal{U}(\tau) = -e^{\beta\Omega}\sum_{n,m}|\langle n|U|m\rangle|^2 e^{-\beta E_n}e^{\tau(E_n-E_m)} \tag{3.108}$$

The frequency transform is

$$\mathcal{U}(i\omega_n) = -e^{\beta\Omega}\sum_{n,m}|\langle n|U|m\rangle|^2 e^{-\beta E_n}\int_0^\beta d\tau e^{i\omega_n\tau}e^{\tau(E_n-E_m)}|$$
$$= e^{\beta\Omega}\sum_{m,n}|\langle n|U|m\rangle|^2\frac{e^{-\beta E_n} - e^{-\beta E_m}}{i\omega_n + E_n - E_m} \tag{3.109}$$

where again $\exp(\beta i\omega_n) = 1$ for bosons. This result should be compared with the retarded function in (3.104). They differ only in the frequencies in the energy denominator since the Matsubara result has $i\omega_n$ where the retarded function has $\omega + i\delta$. The Matsubara function can be changed to a retarded one with just this alteration:

$$\underset{i\omega_n \to \omega + i\delta}{\text{change}}\ \mathcal{U}(i\omega_n) = \bar{U}_{\text{ret}}(\omega) \tag{3.110}$$

This step is called an analytic continuation. The same method can be used to show the same identity for the other Green's functions:

$$\underset{i\omega_n \to \omega + i\delta}{\text{change}} \mathcal{G}(\mathbf{p}, i\omega_n) = G_{\text{ret}}(\mathbf{p}, \omega) \tag{3.111}$$

$$\underset{i\omega_n \to \omega + i\delta}{\text{change}} \mathcal{D}(\mathbf{q}, i\omega_n) = D_{\text{ret}}(\mathbf{q}, \omega) \tag{3.112}$$

This relationship to the retarded function is one of the primary reasons that the Matsubara functions are so useful. After they are evaluated, this simple analytical continuation then yields the retarded function, which is the function of physical interest. The advanced functions are obtained by the analytic continuation $i\omega_n \to \omega - i\delta$, which is obviously true since the advanced function is the complex conjugate of the retarded function.

Another quantity of great importance is the *spectral function*, which is also called the *spectral density function*. It is the imaginary part of any retarded function multiplied by 2; for example,

$$R(\omega) = -2\text{Im}[\bar{U}_{\text{ret}}(\omega)] \tag{3.113}$$

$$B(\mathbf{q}, \omega) = -2\text{Im}[D_{\text{ret}}(\mathbf{q}, \omega)] \tag{3.114}$$

$$A(\mathbf{p}, \omega) = -2\text{Im}[G_{\text{ret}}(\mathbf{p}, \omega)] \tag{3.115}$$

There is not a formal symbol for this quantity, so several are used. From the representation of the retarded function (3.104), the only complex part is

$$\frac{1}{\omega + E_n - E_m + i\delta} = P\frac{1}{\omega + E_n - E_m} - i\pi\delta(\omega + E_n - E_m)$$

which gives for $R(\omega)$

$$R(\omega) = e^{\beta\Omega} \sum_{m,n} |\langle n|U|m\rangle|^2 \{e^{-\beta E_n} - e^{-\beta E_m}\} 2\pi\delta(\omega + E_n - E_m)$$

The temperature factors can be regrouped to give

$$e^{-\beta E_n}\{1 - e^{-\beta(E_m - E_n)}\} \tag{3.116}$$

so that

$$R(\omega) = 2\pi(1 - e^{-\beta\omega})e^{\beta\Omega} \sum_{m,n} |(n|U|m\rangle|^2 \delta(\omega + E_n - E_m) \tag{3.117}$$

It is now possible to write the retarded or Matsubara functions as integrals over these expressions:

$$\bar{U}_{\text{ret}}(\omega) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{R(\omega')}{\omega - \omega' + i\delta}$$

$$\mathcal{U}(i\omega_n) = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{R(\omega')}{i\omega_n - \omega'} \tag{3.118}$$

These identities follow directly from the prior results (3.104), (3.109), and (3.117). An expression of this form is called a *Lehmann representation* (1954), and it was first used in quantum electrodynamics.

Much of this book is devoted to studying electrons, and the spectral function for electrons is calculated often in a variety of problems. Here is a good place to discuss some of its general features. The representation for the retarded function is:

$$G_{\text{ret}}(\mathbf{p}, \omega) = e^{\beta\Omega} \sum_{n,m} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 \frac{e^{-\beta E_n} + e^{-\beta E_m}}{\omega + E_n - E_m + i\delta} \tag{3.119}$$

This equation is similar to the boson result (3.104), the only difference is the plus sign in $e^{-\beta E_n} + e^{-\beta E_m}$. The plus sign follows directly from the plus sign between the two terms in the definition (3.82). The spectral function for the electron is then

$$A(\mathbf{p}, \omega) = 2\pi e^{\beta\Omega} \sum_{n,m} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 [e^{-\beta E_n} + e^{-\beta E_m}]\delta(\omega + E_n - E_m) \tag{3.120}$$

This quantity is absolutely positive for all values of the variables $(\mathbf{p}, \omega)$ since the right-hand side of (3.120) contains only positive factors: $A(\mathbf{p}, \omega) > 0$. This positiveness is an important feature, since $A(\mathbf{p}, \omega)$ is interpreted as a probability function. The spectral functions for bosons do not have this property, since they are sometimes plus and sometimes minus. One can show, however, that they are always plus for $\omega > 0$ and always minus for $\omega < 0$, which follows from Eq. (3.117).

Another important feature of the electron spectral function is obtained by integrating over all frequencies:

$$1 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A(\mathbf{p}, \omega) \tag{3.121}$$

This important theorem, actually a sum rule, is proved by integrating the representation (3.120):

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A(\mathbf{p}, \omega) = e^{\beta\Omega} \sum_{n,m} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 [e^{-\beta E_n} + e^{-\beta E_m}] \tag{3.122}$$

This expression can be simplified by eliminating the summations over n and m. This elimination is achieved by reversing the steps by which we derived (3.104). First one relabels n and m in the second term:

$$\begin{aligned} &= e^{\beta\Omega} \sum_{n,m} e^{-\beta E_n} [\langle n|C_{\mathbf{p}\sigma}|m\rangle \langle m|C^\dagger_{\mathbf{p}\sigma}|n\rangle \\ &\quad + \langle n|C^\dagger_{\mathbf{p}\sigma}|m\rangle \langle m|C_{\mathbf{p}\sigma}|n\rangle] \end{aligned} \tag{3.123}$$

Then one can eliminate the summation over m to give

$$= e^{\beta\Omega} \sum_n e^{-\beta E_n} \langle n|[C_{\mathbf{p}\sigma} C^\dagger_{\mathbf{p}\sigma} + C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma}]|n\rangle \tag{3.124}$$

The anticommutator in brackets yields unity, which proves

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A(\mathbf{p}, \omega) = e^{\beta\Omega} \text{Tr}(e^{-\beta K}) = 1 \tag{3.125}$$

The spectral function may be obtained for a noninteracting electron. In the definition (3.82) set $K_0 = H_0 - \mu N$ and obtain

$$C_{\mathbf{p}\sigma}(t) = e^{-it\xi_{\mathbf{p}}} C_{\mathbf{p}\sigma} \tag{3.126}$$

so that

$$G^{(0)}_{\text{ret}}(\mathbf{p}, t-t') = -i\Theta(t-t')e^{-i(t-t')\xi_{\mathbf{p}}}\langle(C_{\mathbf{p}\sigma}C^{\dagger}_{\mathbf{p}\sigma} + C^{\dagger}_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma})\rangle \tag{3.127}$$

$$= -i\Theta(t-t')e^{-i(t-t')\xi_{\mathbf{p}}} \tag{3.128}$$

The Fourier transform is

$$G^{(0)}_{\text{ret}}(\mathbf{p}, E) = \frac{1}{E - \xi_{\mathbf{p}} + i\delta} \tag{3.129}$$

The factor of $i\delta$ was inserted for convergence. It has one sign $\delta > 0$, even in many electron systems with a Fermi surface. The retarded functions do not have $\delta_{\mathbf{p}}$ changing sign at the Fermi surface, which makes them easier to use than the zero-temperature Green's functions introduced in Chapter 2.

The spectral function for the noninteracting Green's function is

$$A^{(0)}(\mathbf{p}, E) = 2\pi\delta(E - \xi_{\mathbf{p}}) \tag{3.130}$$

It is just a delta function. The spectral function $A(\mathbf{p}, E)$ is interpreted as a probability function. It is the probability that an electron has momentum $\mathbf{p}$ and energy E. For a free, or non-interacting, particle then $E = \xi_{\mathbf{p}}$ so the probability distribution is a delta function: Here there is only one value E for each $\xi_{\mathbf{p}}$ and vice versa. Equation (3.130) is plotted graphically in Fig. 3.1. For a fixed value of $\mathbf{p}$, a plot of the spectral function vs. E is a sharp delta function (here given a small width to aid the eye) at $E = \xi_{\mathbf{p}}$. When $A(\mathbf{p}, E)$ is computed for interacting systems, the broad distribution is found as the one labeled A in Fig. 3.1. There is a band of E values for each $\mathbf{p}$, which is not surprising. When the electron scatters, it has a nonzero mean free path, and there is some uncertainty in its momentum or energy or both. So $\mathbf{p}$ and E are treated as separate variables and both are summed over when evaluating physical quantities. The spectral function $A(\mathbf{p}, E)$ appears in these summations and gives the proper probability weighting between these variables.

Another quantity to evaluate, for an interacting electron system, is the number of electrons in a momentum state $\mathbf{p}$, which is

$$n_{\mathbf{p}} = (C^{\dagger}_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}) \tag{3.131}$$

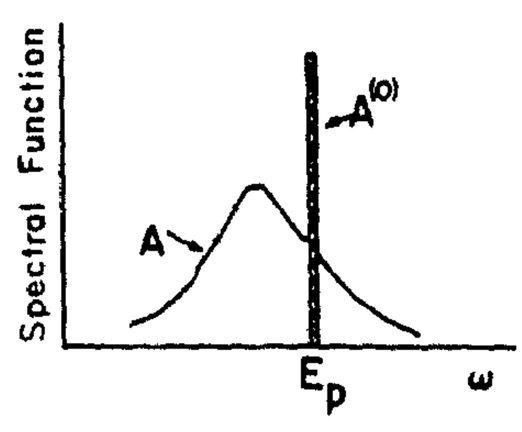


FIGURE 3.1 The noninteracting spectral function $A^{(0)}$ is represented by a delta function, while the interacting spectral function A often has a nonzero width.

For a noninteracting electron system, the number is trivially given by (3.50). In the interacting system, again introduce the representation

$$n_{\mathbf{p}} = e^{\beta\Omega} \sum_{n,m} \langle m|e^{-\beta K} C^{\dagger}_{\mathbf{p}\sigma}|n\rangle \langle n|C_{\mathbf{p}\sigma}|m\rangle \tag{3.132}$$

$$= e^{\beta\Omega} \sum_{n,m} e^{-\beta E_m} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 \tag{3.133}$$

This result $n_{\mathbf{p}}$ should be compared with $A(\mathbf{p}, E)$ in (3.120). The two expressions differ in the factors

$$(e^{-\beta E_n} + e^{-\beta E_m})\delta(E + E_n - E_m) = e^{-\beta E_m}(e^{-\beta(E_n - E_m)} + 1)\delta(E + E_n - E_m)$$
$$= e^{-\beta E_m}(e^{\beta E} + 1)\delta(E + E_n - E_m)$$

Consider the integral using (3.120)

$$\int_{-\infty}^{\infty} \frac{dE}{2\pi} \frac{1}{e^{\beta E} + 1} A(\mathbf{p}, E) = e^{\beta\Omega} \sum_{n,m} e^{-\beta E_m} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 = n_{\mathbf{p}} \tag{3.134}$$

$$n_{\mathbf{p}} = \int_{-\infty}^{\infty} \frac{dE}{2\pi} n_F(E) A(\mathbf{p}, E) \tag{3.135}$$

The factor $n_F(E)$ is the fermion occupation factor at nonzero temperatures. Remember that the energy E is measured with respect to the chemical potential. The thermal occupation probability is determined by the energy E of the particles. The number of particles in a state $\mathbf{p}$ is obtained by summing over all energies E, weighted by the spectral function which gives the probability that a particle in state $\mathbf{p}$ has energy E, and also by multiplying by the thermal occupation factor $n_F(E)$. This reasonable expression provides further examples of the use of the spectral function. In the limit of no interactions, then $A(\mathbf{p}, E)$ becomes the noninteracting spectral function (3.130), and Eq. (3.50) is recovered again.

For phonons, the average number of phonons in a state $\mathbf{q}$ is $N_{\mathbf{q}} = \langle a^{\dagger}_{\mathbf{q}} a_{\mathbf{q}}\rangle$, and one can similarly show that

$$2N_{\mathbf{q}} + 1 = \langle A^{\dagger}_{\mathbf{q}} A_{\mathbf{q}}\rangle = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} n_B(\omega) B(\mathbf{q}, \omega) \tag{3.136}$$

The phonon spectral function $B(\mathbf{q}, \omega)$ was defined earlier in (3.114). The factor $n_B(\omega)B(\mathbf{q}, \omega)$ is always positive for phonons and could serve as the temperature-dependent probability of having phonons with $\mathbf{q}$ and ω. The noninteracting phonon spectral function is

$$B^{(0)}(\mathbf{q}, \omega) = 2\pi[\delta(\omega - \omega_{\mathbf{q}}) - \delta(\omega + \omega_{\mathbf{q}})] \tag{3.137}$$

The retarded Green's functions will be calculated from the Matsubara functions. In the next section it is shown that the Matsubara functions have a Dyson equation of the form

$$\mathcal{G}(\mathbf{p}, ip_n) = \frac{1}{ip_n - \xi_{\mathbf{p}} - \Sigma(\mathbf{p}, ip_n)} \tag{3.138}$$

$$\mathcal{D}(\mathbf{q}, i\omega_n) = \frac{-2\omega_{\mathbf{q}}}{\omega_n^2 + \omega_{\mathbf{q}}^2 + 2\omega_{\mathbf{q}} P(\mathbf{q}, i\omega_n)} \tag{3.139}$$

The self-energy functions Σ and P will be calculated according to rules described later. Define the retarded self-energies according to (3.111):

$$\underset{ip_n \to E + i\delta}{\text{change}} \Sigma(\mathbf{p}, ip_n) \to \Sigma_{\text{ret}}(\mathbf{p}, E) = \text{Re}\Sigma_{\text{ret}}(\mathbf{p}, E) + i\text{Im}\Sigma_{\text{ret}}(\mathbf{p}, E)$$

$$\underset{i\omega_n \to \omega + i\delta}{\text{change}} P(\mathbf{q}, i\omega_n) \to P_{\text{ret}}(\mathbf{q}, \omega) \tag{3.140}$$

The retarded Green's function will also have a Dyson equation

$$G_{\text{ret}}(\mathbf{p}, E) = \frac{1}{E + i\delta - \xi_{\mathbf{p}} - \Sigma_{\text{ret}}(\mathbf{p}, E)} \tag{3.141}$$

as derived from (3.111). The spectral function for the electron is rewritten in terms of the retarded self-energies,

$$A(\mathbf{p}, E) = \frac{-2\text{Im}\Sigma_{\text{ret}}(\mathbf{p}, E)}{[E - \xi_{\mathbf{p}} - \text{Re}\Sigma_{\text{ret}}(\mathbf{p}, E)]^2 + [\text{Im}\Sigma_{\text{ret}}(\mathbf{p}, E)]^2} \tag{3.142}$$

where $\text{Im}\Sigma < 0$ so that $A > 0$. One way to obtain the spectral function is to evaluate the self-energies. The method for doing this is described in the next section.

The formal distinctions between the Matsubara, retarded, and advanced Green's functions are best understood by some simple examples. There are some simple functions which have the correct analytical properties. For example, consider a self-energy operator which has the following functional form,

$$\Sigma(\mathbf{p}, Z) = C \ln[f(\mathbf{p}) - Z] \tag{3.143}$$

where Z is a complex variable representing the frequency. Take C as a constant and $f(\mathbf{p})$ as some function of momentum. The Matsubara self-energy is evaluated at the points ip_n:

$$\Sigma(\mathbf{p}, ip_n) = C \ln[f(\mathbf{p}) - ip_n] \tag{3.144}$$

The analytic continuation $ip_n \to E \pm i\delta$ to the real axis has the following values. For the retarded function, $ip_n \to E + i\delta$

$$\Sigma_{\text{ret}}(\mathbf{p}, E) = C \ \ln|f(\mathbf{p}) - E| - i\pi C\Theta[E - f(\mathbf{p})] \tag{3.145}$$

while for the advanced function, $ip_n \to E - i\delta$,

$$\Sigma_{\text{adv}}(\mathbf{p}, E) = C \ \ln|f(\mathbf{p}) - E| + i\pi C\Theta[E - f(\mathbf{p})] \tag{3.146}$$

These two self-energies differ in the region $E > f(\mathbf{p})$, because their imaginary parts have the opposite sign. This difference agrees with our general theorem that

$$G_{\text{ret}}(\mathbf{p}, E)^* = G_{\text{adv}}(\mathbf{p}, E) \tag{3.147}$$

which also implies that

$$\Sigma_{\text{ret}}(\mathbf{p}, E)^* = \Sigma_{\text{adv}}(\mathbf{p}, E) \tag{3.148}$$

This functional behavior is shown in Fig. 3.2. There is a branch cut on the real axis for $E > f(\mathbf{p})$. This branch cut just expresses the fact that $\ln(f - Z)$ is not a continuous function of Z across the real axis for $E > f$, since the imaginary part changes sign.

Another example which has similar analytical properties is

$$\Sigma(\mathbf{p}, Z) = C[f(\mathbf{p}) - Z]^{1/2} \tag{3.149}$$

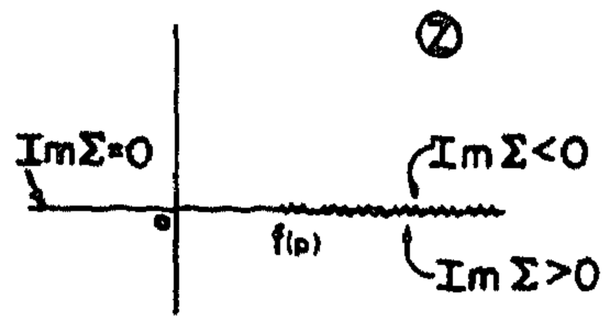


FIGURE 3.2

This function also has a branch cut for $E > f(\mathbf{p})$, with $\operatorname{Im}\Sigma < 0$ above the branch cut and $\operatorname{Im}\Sigma > 0$ below. In fact, a branch cut is a necessary feature whenever $\operatorname{Im}\Sigma \neq 0$. Whenever $\operatorname{Im}\Sigma \neq 0$, then

$$\Sigma(\mathbf{p}, E + i\delta) \neq \Sigma(\mathbf{p}, E - i\delta) \tag{3.150}$$

which requires a branch cut in the analytic function. When self-energy functions are evaluated, they are often given by logarithmic or square root functions.

When a branch cut occurs and $\operatorname{Im}\Sigma \neq 0$, then the spectral function is given by (3.142). In frequency regions where $\operatorname{Im}\Sigma = 0$ and there is no branch cut, then take the limit of $\operatorname{Im}\Sigma \to 0$ and obtain

$$\operatorname{Im}\Sigma = 0: \quad A(\mathbf{p}, E) = 2\pi\delta(E - \xi_{\mathbf{p}} - \operatorname{Re}\Sigma_{\text{ret}}(\mathbf{p}, E)) \tag{3.151}$$

Here the spectral function is again a delta function, but now the real part of the self-energy may be nonzero, and usually is, so that it affects the spectral function. Denote by $E(\mathbf{p})$ the solution to the equation

$$E(\mathbf{p}) - \mu = \xi_{\mathbf{p}} + \operatorname{Re}\Sigma_{\text{ret}}(\mathbf{p}, E(\mathbf{p}) - \mu) \tag{3.152}$$

Assume that there is a problem in which (3.152) is satisfied when $\operatorname{Im}\Sigma = 0$. Recall that if $g(x) = 0$ at $x = x_0$, then delta functions have the property

$$\delta[g(x)] = \frac{\delta(x - x_0)}{|g'(x_0)|} \tag{3.153}$$

and the spectral function is written as

$$A(\mathbf{p}, E) = 2\pi Z(\mathbf{p})\delta(E - E(\mathbf{p}) + \mu) \tag{3.154}$$

$$Z(\mathbf{p}) = \frac{1}{\left|1 - \dfrac{\partial}{\partial E}\Sigma_{\text{ret}}(\mathbf{p}, E)\right|_{E=E(\mathbf{p})-\mu}} \tag{3.155}$$

The factor $Z(\mathbf{p})$ is called a *renormalization factor*. If we recall the facts that (1) $A(\mathbf{p}, E) > 0$ and (2) $\int dE/(2\pi)A(\mathbf{p}, E) = 1$, then it is easy to show that $Z(\mathbf{p}) < 1$. The strength of the delta function peak is always less than or equal to unity. An example is illustrated in Fig. 3.3. The

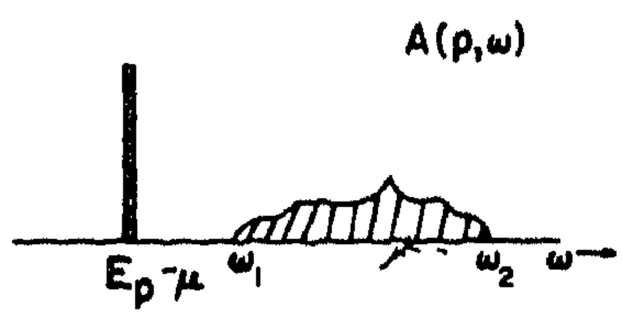


FIGURE 3.3 The spectral function $A(\mathbf{p}, \omega)$ will have a delta function peak whenever $\operatorname{Im}\Sigma = 0$, which is shown here at $\omega = E_p - \mu$. The spectral function also has values in regions where $\operatorname{Im}\Sigma \neq 0$, as shown by the crosshatched distribution.

spectral function is shown with a sharp delta function peak at $\omega = E_p - \mu$ and a continuous spectra in the region $\omega_1 < \omega < \omega_2$. The latter region is where $\mathrm{Im}\,\Sigma \neq 0$. Since the total integrated area under the entire spectra is unity, the existence of any areas where $\mathrm{Im}\,\Sigma \neq 0$ implies that the renormalization factor $Z(\mathbf{p})$ for the delta function is less than unity.

Equation (3.152) may be used to define the effective mass. Assume that the noninteracting states are free particles, so that

$$\xi_{\mathbf{p}} = \frac{p^2}{2m} - \mu = \varepsilon_p - \mu \tag{3.156}$$

Furthermore, assume at low momentum that $E_p \equiv E(\mathbf{p})$ varies quadratically with momentum,

$$E_p = E_0 + \frac{p^2}{2m^*} + O(p^4) \tag{3.157}$$

The proportionality constant is the inverse effective mass m^*

$$\frac{m}{m^*} = \frac{\partial E_p}{\partial \varepsilon_p} \tag{3.158}$$

and the derivative of (3.152) is

$$\frac{\partial E_p}{\partial \varepsilon_p} = \lim_{\varepsilon_p \to 0}\left(1 + \frac{\partial}{\partial \varepsilon_p}\mathrm{Re}\,\Sigma(\mathbf{p}, E) + \left[\frac{\partial}{\partial E}\mathrm{Re}\,\Sigma(\mathbf{p}, E)\right]\frac{\partial E_p}{\partial \varepsilon_p}\right) \tag{3.159}$$

The last term on the right contains the factor (3.158), which is brought to the left to finally obtain

$$\frac{m}{m^*} = \lim_{\varepsilon_p \to 0}\left[\frac{1 + (\partial/\partial\varepsilon_p)\mathrm{Re}\,\Sigma(\mathbf{p}, E_0 - \mu)}{1 - (\partial/\partial E_0)\mathrm{Re}\,\Sigma(\mathbf{p}, E_0 - \mu)}\right] \tag{3.160}$$

This formula will be used frequently to obtain the effective mass from self-energy calculations.

3.4. DYSON'S EQUATION

The Matsubara Green's functions are evaluated by the same method of Feynman diagram techniques that were introduced in Chapter 2 for the case of zero temperature. These methods are slightly modified to account for complex times and frequencies, but otherwise the S-matrix expansion appears to be very similar. The ideas behind expanding the S matrix are presented, and Dyson's equation is rederived for the Matsubara Green's functions.

Consider the case of the electron Green's function:

$$\mathcal{G}(\mathbf{p}, \tau) = -e^{\beta\Omega}\mathrm{Tr}[e^{-\beta K}T_\tau(e^{\tau K}C_{\mathbf{p}\sigma}e^{-\tau K})C^\dagger_{\mathbf{p}\sigma}] \tag{3.161}$$

$$e^{-\beta\Omega} = \mathrm{Tr}(e^{-\beta K}) \tag{3.162}$$

Again consider the general case where

$$K = K_0 + V = H_0 - \mu N + V \tag{3.163}$$

$$H = H_0 + V \tag{3.164}$$

where K_0 is a problem which can be solved, so that we know its complete set of states. The Hamiltonians under consideration usually have the property that they commute with the number operator:

$$[H_0, N] = 0 \tag{3.165}$$

$$[H, N] = 0 \tag{3.166}$$

In this case, simultaneous eigenstates of H_0 and N can be defined and also of H and N. Call H_0 or K_0 the noninteracting problem, and V is the perturbation whose effects are being evaluated. Consider the operators in the interaction representation

$$U(\tau) = e^{\tau K_0} e^{-\tau K} \tag{3.167}$$

$$U^{-1}(\tau) = e^{\tau K} e^{-\tau K_0} \tag{3.168}$$

A caret is put above operators to denote their τ dependence with respect to the noninteracting operators:

$$\hat{C}_{\mathbf{p}\sigma}(\tau) = e^{\tau K_0} C_{\mathbf{p}\sigma} e^{-\tau K_0} = e^{-\tau \xi_{\mathbf{p}}} C_{\mathbf{p}\sigma} \tag{3.169}$$

The Green's function in (3.161) is written for $\tau > 0$ as

$$\begin{aligned} \mathcal{G}(\mathbf{p}, \tau) &= -e^{\beta\Omega} \mathrm{Tr}[e^{-\beta K_0}(e^{\beta K_0} e^{-\beta K})(e^{\tau K} e^{-\tau K_0})(e^{\tau K_0} C_{\mathbf{p}\sigma} e^{-\tau K_0}) \\ &\quad \times (e^{\tau K_0} e^{-\tau K}) C^\dagger_{\mathbf{p}\sigma}] \\ &= -\frac{\mathrm{Tr}(e^{-\beta K_0} U(\beta) U^{-1}(\tau) \hat{C}_{\mathbf{p}\sigma}(\tau) U(\tau) C^\dagger_{\mathbf{p}\sigma}]}{\mathrm{Tr}[e^{-\beta K_0} U(\beta)]} \end{aligned} \tag{3.170}$$

The thermodynamic potential in the denominator has been replaced by the equivalent factor

$$\begin{aligned} e^{-\beta\Omega} &= \mathrm{Tr}(e^{-\beta K}) = \mathrm{Tr}[e^{-\beta K_0}(e^{\beta K_0} e^{-\beta K})] \\ &= \mathrm{Tr}[e^{-\beta K_0} U(\beta)] \end{aligned} \tag{3.171}$$

A similar substitution has been made on the factor of $\exp(-\beta K)$ in the numerator.

The operator $U(\tau)$ can be solved in terms of τ-ordered products. Consider the derivative

$$\frac{\partial}{\partial \tau} U(\tau) = e^{\tau K_0}(K_0 - K) e^{-\tau K} = -e^{\tau K_0} V e^{-\tau K} \tag{3.172}$$

which can be expressed in the interaction representation as

$$\frac{\partial}{\partial \tau} U(\tau) = -(e^{\tau K_0} V e^{-\tau K_0})(e^{\tau K_0} e^{-\tau K}) = -\hat{V}(\tau) U(\tau) \tag{3.173}$$

This equation for $U(\tau)$ may be solved, at least formally, by repeated integrations and using $U(0) = 1$:

$$U(\tau) = 1 - \int_0^\tau d\tau_1 \hat{V}(\tau_1) U(\tau_1) \tag{3.174}$$

$$\begin{aligned} &= 1 - \int_0^\tau d\tau_1 \hat{V}(\tau_1) + (-1)^2 \int_0^\tau d\tau_1 \int_0^{\tau_1} d\tau_2 \hat{V}(\tau_1) \hat{V}(\tau_2) U(\tau_2) \\ &= \sum_{n=0}^{\infty} (-1)^n \int_0^\tau d\tau_1 \cdots \int_0^{\tau_{n-1}} d\tau_n \hat{V}(\tau_1) \cdots \hat{V}(\tau_n) \end{aligned} \tag{3.175}$$

The summation can be expressed as an ordered product,

$$U(\tau) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^\tau d\tau_1 \cdots \int_0^\tau d\tau_n [T_\tau \hat{V}(\tau_1) \cdots \hat{V}(\tau_n)] \tag{3.176}$$

$$= T_\tau \exp\left[-\int_0^\tau d\tau_1 \hat{V}(\tau_1)\right] \tag{3.177}$$

and finally as a τ-ordered exponential integral. The strict definition of (3.177) is just (3.176). Next consider the definition

$$S(\tau_1, \tau_2) = T_\tau \exp\left[-\int_{\tau_1}^{\tau_2} d\tau_1 \hat{V}(\tau_1)\right] \tag{3.178}$$

$$S(\tau) = U(\tau) \tag{3.179}$$

It is easy to prove the following operator identities:

$$S(\tau_2, \tau_1) = U(\tau_2) U^{-1}(\tau_1) \tag{3.180}$$

$$S(\tau_3, \tau_2) S(\tau_2, \tau_1) = S(\tau_3, \tau_1) \tag{3.181}$$

In this notation the Green's function (3.170) may be rewritten for $\tau > 0$ as

$$\mathcal{G}(\mathbf{p}, \tau) = -\frac{\mathrm{Tr}[e^{-\beta K_0} T_\tau S(\beta, \tau) \hat{C}_{\mathbf{p}\sigma}(\tau) S(\tau) \hat{C}^\dagger_{\mathbf{p}\sigma}(0)]}{\mathrm{Tr}[e^{-\beta K_0} S(\beta)]} \tag{3.182}$$

Using the above properties of $S(\beta, \tau)$ and taking advantage of the freedom to rearrange terms within the τ-ordering operator, the numerator may be expressed as

$$\mathcal{G}(\mathbf{p}, \tau) = -\frac{\mathrm{Tr}[e^{-\beta K_0} T_\tau S(\beta) \hat{C}_{\mathbf{p}\sigma}(\tau) \hat{C}^\dagger_{\mathbf{p}\sigma}(0)]}{\mathrm{Tr}[e^{-\beta K_0} S(\beta)]} \tag{3.183}$$

From now on the trace over $\exp(-\beta K_0)$ is written with a subscript 0,

$$\mathrm{Tr}[e^{-\beta K_0}] = {}_0\langle \theta \rangle \tag{3.184}$$

where θ is any operator. The Green's function is

$$\mathcal{G}(\mathbf{p}, \tau) = -\frac{{}_0\langle T_\tau S(\beta) \hat{C}_{\mathbf{p}\sigma}(\tau) \hat{C}^\dagger_{\mathbf{p}\sigma}(0) \rangle}{{}_0\langle S(\beta) \rangle} \tag{3.185}$$

We continue to write a caret over $\hat{C}^\dagger_{\mathbf{p}\sigma}(0)$. Although it has no τ dependence when $\tau = 0$, the $\tau = 0$ label is important for ordering the operator. Equation (3.183) has been proved only for the case $\tau > 0$, but this case is the only one of interest. The Tr operator in (3.183) really means (3.182). The S matrix is divided into parts, which are ordered with respect to the other operators $\hat{C}_{\mathbf{p}\sigma}(\tau)$ and $\hat{C}^\dagger_{\mathbf{p}\sigma}(0)$. This form of the Green's function is similar to the zero-temperature result (2.50).

The Green's function is evaluated, at least formally, by expanding the S matrix in the numerator:

$$\begin{aligned} &{}_0\langle T_\tau S(\beta) \hat{C}_{\mathbf{p}\sigma}(\tau) \hat{C}^\dagger_{\mathbf{p}\sigma}(0) \rangle \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_n \, {}_0\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau) \hat{V}(\tau_1) \cdots \hat{V}(\tau_n) \hat{C}^\dagger_{\mathbf{p}\sigma}(0) \rangle \end{aligned} \tag{3.186}$$

Each of the nth terms are evaluated by applying Wick's theorem to the brackets and thereby expressing the brackets as combinations of the non-interacting Green's functions $\mathcal{G}^{(0)}$ and $D^{(0)}$. Wick's theorem is applied as before to give, for example,

$$
\begin{aligned}
&{}_0\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau)\hat{C}^\dagger_{\mathbf{k}s}(\tau_1)\hat{C}_{\mathbf{k}'s}(\tau_1)\hat{C}^\dagger_{\mathbf{p}\sigma}(0)]\rangle \\
&\quad = \delta_{\mathbf{p}=\mathbf{k}}\delta_{\mathbf{p}=\mathbf{k}'}\delta_{s=\sigma}\mathcal{G}^{(0)}(\mathbf{p},\tau-\tau_1)\mathcal{G}^{(0)}(\mathbf{p},\tau_1) - \delta_{\mathbf{k}=\mathbf{k}'}n_{\mathbf{k}}\mathcal{G}^{(0)}(\mathbf{p},\tau)
\end{aligned}
\tag{3.187}
$$

Wick's theorem has a double meaning for Green's functions at nonzero temperatures. The first meaning is just the pairing feature: start with a state $|n\rangle$ (in the trace) and operate on this state by some C_λ. In order to get back to $\langle n|$ at the other side of the trace, there must be another operator of the form $C^\dagger_\lambda$ appearing in the product of operators. The pairing was used in Wick's theorem for zero temperatures, and the same use is made here. Whenever there is a bracket with M creation and M destruction operators, each of the creation operators must correspond to the same state as one of the destruction operators and vice versa.

The other feature of Wick's theorem concerns the thermodynamic average. The brackets ${}_0\langle\cdots\rangle$ mean the average in (3.184), and usually

$$
{}_0\langle AB\rangle \neq {}_0\langle A\rangle_0\langle B\rangle \tag{3.188}
$$

where A and B are two arbitrary operators. Usually, the product of averages is not equal to the average of products. However, in Wick's theorem, we do just that: a product of operators is paired up, and each pair is thermodynamically averaged separately. This procedure does give the right answer as long as H_0 is only bilinear in the operators. Any errors that are made vanish in the limit of infinite volume. One exception is for macroscopic quantum states as occur in superfluids. As an example, consider the evaluation of

$$
W = \left\langle T_\tau\left[\sum_{\mathbf{p},s} C^\dagger_{\mathbf{p},s}(\tau)C_{\mathbf{p},s}(\tau)\sum_{\mathbf{k},s'} C^\dagger_{\mathbf{k},s'}(\tau')C_{\mathbf{k},s'}(\tau')\right]\right\rangle \tag{3.189}
$$

where the operators refer to fermions. First, do it exactly. The combination

$$
{}_0\langle C^\dagger_{\mathbf{p},s}(\tau)C_{\mathbf{p},s}(\tau)\rangle = n_{\mathbf{p}} \tag{3.190}
$$

is the number operator. It is assumed to commute with the Hamiltonian and has no τ dependence. In the thermodynamic average, the cases of $\mathbf{p}=\mathbf{k}, s'=s$ must be treated separately from $\mathbf{p}\neq\mathbf{k}$ or $s'\neq s$. The occupation numbers are assumed to be independent of the spin index:

$$
\begin{aligned}
W &= \sum_{s,s'}\sum_{\mathbf{p}\neq\mathbf{k}} {}_0\langle C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}C^\dagger_{\mathbf{k},s'}C_{\mathbf{k},s'}\rangle + \sum_{\mathbf{p},s}[{}_0\langle C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}\rangle \\
&\quad + {}_0\langle C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}C^\dagger_{\mathbf{p},-s}C_{\mathbf{p},-s}\rangle] && (3.191)\\
&= \sum_{s,s'}\sum_{\mathbf{p}\neq\mathbf{k}} n_{\mathbf{p}}n_{\mathbf{k}} + \sum_{\mathbf{p},s}[n_{\mathbf{p}} + n^2_{\mathbf{p}}] && (3.192)\\
&= \left(\sum_{\mathbf{p},s} n_{\mathbf{p}}\right)^2 + \sum_{\mathbf{p},s} n_{\mathbf{p}}[1-n_{\mathbf{p}}] && (3.193)
\end{aligned}
$$

The average of ${}_0\langle C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}\rangle$ gives only one factor of $n_{\mathbf{p}}$ since the occupation number $C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}$ is either zero or one. Squaring it still gives zero or one. However, the average ${}_0\langle C^\dagger_{\mathbf{p},s}C_{\mathbf{p},s}C^\dagger_{\mathbf{p},-s}C_{\mathbf{p},-s}\rangle$ gives $n^2_{\mathbf{p}}$ since the two spin states are averaged independently.

The same problem can be solved by the application of Wick's theorem which gives

$$W = \left(\sum_{\mathbf{p},s} n_{\mathbf{p}}\right)^2 - \sum_{\mathbf{p},s} \mathcal{G}^{(0)}(\mathbf{p}, \tau - \tau')\mathcal{G}^{(0)}(\mathbf{p}, \tau' - \tau) \tag{3.194}$$

Recall that $\mathcal{G}^{(0)}(\mathbf{p}, \tau) = -[\Theta(\tau) - n_{\mathbf{p}}]\exp(-\xi_{\mathbf{p}}\tau)$. Then the above combination of Green's functions gives ($\tau' = 0$)

$$\mathcal{G}^{(0)}(\mathbf{p}, \tau)\mathcal{G}^{(0)}(\mathbf{p}, -\tau) = [\Theta(\tau) - n_{\mathbf{p}}][\Theta(-\tau) - n_{\mathbf{p}}] = -n_{\mathbf{p}}[1 - n_{\mathbf{p}}]$$

and this method yields the same answer (3.193) as the exact method.

Diagrams in the S-matrix expansion are classified as connected or disconnected according to the conventions discussed in Sec. 2.6. Only connected diagrams are retained, since the disconnected diagrams are canceled by the vacuum polarization diagrams. For Matsubara Green's functions, the vacuum polarization terms come from the denominator:

$$e^{-\beta\Omega} = {}_0\langle S(\beta)\rangle = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_{n0}\langle T_\tau \hat{V}(\tau_1) \cdots \hat{V}(\tau)\rangle \tag{3.195}$$

The expansion of this quantity produces a series of diagrams which just cancel the disconnected parts of the expansion in (3.186). In the zero-temperature case of Chapter 2, it was found that each vacuum polarization graph gave infinity, which was perhaps not a worry since they canceled out of the answer. In the Matsubara formalism, the vacuum polarization terms are all nonzero. The "time" integrals, which are now τ integrals, are only over the range $0 < \tau < \beta$ and the extra integration does not diverge. The Matsubara formalism eliminates this one divergence, which is another small advantage of the method.

The summation in (3.195) actually evaluates a useful quantity: the thermodynamic potential Ω. One might actually wish to calculate it. The method of evaluation is presented in Sec. 3.6, and is called a linked cluster expansion.

The Matsubara Green's function can be reduced to an evaluation of all connected, different diagrams:

$$\mathcal{G}(\mathbf{p}, \tau) = -\sum_{n=0}^{\infty}(-1)^n \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_{n0}\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau)\hat{V}(\tau_1) \cdots \hat{V}(\tau_n)\hat{C}^\dagger_{\mathbf{p}\sigma}(0)\rangle \tag{3.196}$$

where only the different, disconnected diagrams are evaluated. Each diagram is evaluated as a function of τ, and the Green's function of energy is found from the Fourier transform:

$$\mathcal{G}(\mathbf{p}, ip_n) = \int_0^\beta d\tau e^{ip_n\tau}\mathcal{G}(\mathbf{p}, \tau) \tag{3.197}$$

and $p_n = (2n + 1)\pi k_B T$. The terms in the series (3.196) yield self-energy diagrams, which may be collected into Dyson equations:

$$\mathcal{G}(\mathbf{p}, ip_n) = \frac{\mathcal{G}^{(0)}(\mathbf{p}, ip_n)}{1 - \mathcal{G}^{(0)}(\mathbf{p}, ip_n)\Sigma(\mathbf{p}, ip_n)} \tag{3.198}$$

$$\mathcal{D}(\mathbf{q}, i\omega_n) = \frac{\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n)}{1 - \mathcal{D}^{(0)}(\mathbf{q}, i\omega_n)\Pi(\mathbf{q}, i\omega_n)} \tag{3.199}$$

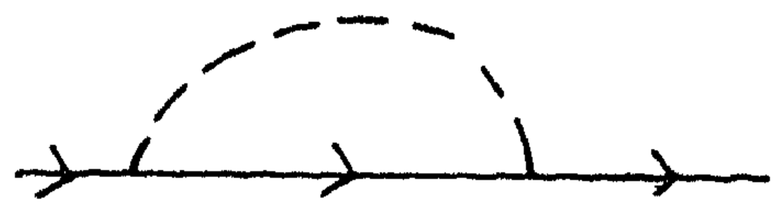


FIGURE 3.4

As an example, consider the basic electron self-energy contribution from the electron–phonon interaction shown in Fig. 3.4. This self-energy was evaluated in (2.111) using zero-temperature methods. Here it is evaluated again using Matsubara methods. The interaction is

$$V = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\mathbf{q}} M_{\mathbf{q}} A_{\mathbf{q}} \sum_{\mathbf{p}\sigma} C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{3.200}$$

The $n = 2$ term in the S-matrix expansion is

$$\begin{aligned} \mathcal{G}_2 = -\frac{1}{\mathcal{V}} \sum_{\mathbf{q}\mathbf{q}'} M_{\mathbf{q}} M_{\mathbf{q}'} \int_0^{\beta} d\tau_1 \int_0^{\beta} d\tau_2 \, {}_0\langle T_\tau \hat{A}_{\mathbf{q}}(\tau_1) \hat{A}_{\mathbf{q}'}(\tau_2)\rangle \\ \times \sum_{\mathbf{k},\mathbf{k}',s,s'} {}_0\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau) \hat{C}^{\dagger}_{\mathbf{k}+\mathbf{q},s}(\tau_1) \hat{C}_{\mathbf{k},s}(\tau_1) \hat{C}^{\dagger}_{\mathbf{k}'+\mathbf{q}',s'}(\tau_2) \\ \times \hat{C}_{\mathbf{k}',s'}(\tau_2) \hat{C}^{\dagger}_{\mathbf{p}\sigma}\rangle \end{aligned} \tag{3.201}$$

The factor of

$${}_0\langle T_\tau \hat{A}_{\mathbf{q}}(\tau_1) \hat{A}_{\mathbf{q}'}(\tau_2)\rangle = -\delta_{\mathbf{q}=-\mathbf{q}'} \mathcal{D}^{(0)}(\mathbf{q}, \tau_1 - \tau_2) \tag{3.202}$$

gives the phonon Green's function. Applying Wick's theorem to the electron bracket yields six terms, but there is only one different disconnected term. It gives the contribution:

$$\begin{aligned} \mathcal{G}_2(\mathbf{p}, \tau) = -\frac{1}{\mathcal{V}} \sum_{\mathbf{q}\mathbf{q}'} M_{\mathbf{q}}^2 \int_0^{\beta} d\tau_1 \int_0^{\beta} d\tau_2 \mathcal{D}^{(0)}(\mathbf{q}, \tau_1 - \tau_2) \\ \times \mathcal{G}^{(0)}(\mathbf{p}, \tau - \tau_1) \mathcal{G}^{(0)}(\mathbf{p} + \mathbf{q}, \tau_1 - \tau_2) \mathcal{G}^{(0)}(\mathbf{p}, \tau_2) \end{aligned} \tag{3.203}$$

The τ integrals may be performed easily, since the τ dependence of the non-interacting Green's functions was derived earlier in (3.49) and (3.72). However, this step is rarely performed. Instead, it is easier to take the Fourier transform in (3.197). The τ integrals are done by using the expansion given in (3.39) and (3.65), such as

$$\mathcal{D}^{(0)}(\mathbf{q}, \tau_1 - \tau_2) = \frac{1}{\beta} \sum_{\omega_n} e^{-i\omega_n(\tau_1 - \tau_2)} \mathcal{D}^{(0)}(\mathbf{q}, i\omega_n) \tag{3.204}$$

$$\mathcal{G}^{(0)}(\mathbf{p}, \tau - \tau_1) = \frac{1}{\beta} \sum_{p_m} e^{-ip_m(\tau - \tau_1)} \mathcal{G}^{(0)}(\mathbf{p}, ip_m) \tag{3.205}$$

so that

$$\mathcal{G}(\mathbf{p}, ip_n) = -\frac{1}{\nu\beta^4}\sum_{\mathbf{q}} M_{\mathbf{q}}^2 \sum_{m,n',n'',n'''} \mathcal{D}^{(0)}(\mathbf{q}, i\omega_m)\mathcal{G}^{(0)}(\mathbf{p}, ip_{n'})$$
$$\times \mathcal{G}^{(0)}(\mathbf{p}-\mathbf{q}, ip_{n''})\mathcal{G}^{(0)}(\mathbf{p}, ip_{n'''}) \int_0^\beta d\tau \int_0^\beta d\tau_1 \int_0^\beta d\tau_2$$
$$\times \exp[ip_n\tau - i\omega_m(\tau_1-\tau_2) - ip_{n'}(\tau-\tau_1)$$
$$- ip_{n''}(\tau_1-\tau_2) - ip_{n'''}\tau_2] \tag{3.206}$$

All three τ integrals are of the form

$$\frac{1}{\beta}\int_0^\beta d\tau e^{i\tau(p_n-p_m)} = \frac{1}{i\beta(p_n-p_m)}(e^{i\beta(p_n-p_m)}-1)$$
$$= \frac{1}{2i\pi(n-m)}(e^{2i\pi(n-m)}-1) = \delta_{nm} \tag{3.207}$$

which are written in the shorthand notation

$$\frac{1}{\beta}\int_0^\beta d\tau e^{i\tau(p_n-p_{n'})} = \delta_{p_n=p_{n'}} \tag{3.208}$$

$$\frac{1}{\beta}\int_0^\beta d\tau e^{i\tau_1(p_{n'}-p_{n''}-\omega_m)} = \delta_{p_{n'}=p_{n''}+\omega_m} \tag{3.209}$$

$$\frac{1}{\beta}\int_0^\beta d\tau e^{i\tau_2(p_{n''}-p_{n'''}+\omega_m)} = \delta_{p_{n''}=p_{n'''}-\omega_m} \tag{3.210}$$

Combining these results yields the final result

$$\mathcal{G}_2(\mathbf{p}, ip_n) = \mathcal{G}^{(0)}(\mathbf{p}, ip_n)^2\Sigma^{(1)}(\mathbf{p}, ip_n) \tag{3.211}$$

where the one-phonon self-energy term is now

$$\Sigma^{(1)}(\mathbf{p}, ip_n) = -\frac{1}{\beta\nu}\sum_{\mathbf{q}\omega_m} M_{\mathbf{q}}^2\mathcal{D}^{(0)}(\mathbf{q}, i\omega_m)\mathcal{G}^{(0)}(\mathbf{p}-\mathbf{q}, ip_n - i\omega_m) \tag{3.212}$$

The self-energy (3.212) in the Matsubara notation is similar to (2.111). The only difference is the use of complex frequencies rather than real ones. Also note there is frequency conservation in this expression. An electron starts with frequency ip_n, and it emits or absorbs a phonon with frequency $i\omega_m$, so the electron goes to an intermediate state with frequency $ip_n - i\omega_m$. Again the summation is taken over internal variables, in this case the momentum $\mathbf{q}$ and the frequency $i\omega_m$. In general, the complex frequency is conserved at each vertex in the Feynman diagram. This conservation does maintain the oddness of fermion frequencies and the evenness for bosons. For example, the electron in (3.212) has the frequency $ip_n - i\omega_m$. Now ip_n is an odd integer, since it was associated with the original electron line, while $i\omega_m$ is an even integer, since it was associated with the phonon. But an odd plus an even is odd, which preserves the oddness of electron frequencies.

The exact Matsubara Green's function may be obtained by finding the exact self-energies. The exact solution is rarely possible, and usually one is content to write down a few diagrams and evaluate the first few terms in the expansion for $\Sigma(\mathbf{p}, ip_n)$. To obtain these terms, draw the Feynman diagrams, conserve momentum and frequency at each vertex, and then

integrate or sum over internal variables. All that is needed now are some *rules for constructing diagrams*:

1. With each internal electron line, associate a quantity $\mathcal{G}^{(0)}(\mathbf{p}, ip_n)$.
2. With each internal phonon line, associate a quantity $M_{\mathbf{q}}^2 D^{(0)}(\mathbf{q}, i\omega_m)$.
3. With each internal Coulomb line, associate a quantity $v_q = 4\pi e^2/q^2$.
4. Conserve momentum and complex frequency at each vertex. Keep in mind that fermion frequencies are odd integers $(2n+1)\pi k_B T$ and boson frequencies are even integers $2\pi n k_B T$. Their oddness and evenness will be maintained in the energy conservation.
5. Sum over internal degrees of freedom: momentum and frequency. Internal variables are all those except the $(\mathbf{p}, ip_n)$ of the self-energy.
6. Multiply the expression by

$$\frac{(-1)^{m+F}(2S+1)^F}{(\beta\nu)^m} \tag{3.213}$$

where F is the number of closed Fermion loops. The $(2S+1)$ factor is a summation over spin degrees of freedom, and $2S+1=2$ for electrons. The integer m is the order of the diagram, as defined earlier in Sec. 2.8.

Several examples are presented. The basic fermion loop in Fig. 3.5(a) occurs in the phonon self-energy and also the Coulomb self-energy. It is

$$P^{(1)}(\mathbf{q}, i\omega_n) = \frac{2}{\beta\nu}\sum_{\mathbf{q}, ip_m} \mathcal{G}^{(0)}(\mathbf{p}, ip_m)\mathcal{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip_m + i\omega_n) \tag{3.214}$$

In Fig. 3.5(b) a phonon line has been added between the two electron lines, which is called a vertex correction. This diagram gives, according to the rules,

$$\begin{aligned} P^{(2)}(\mathbf{q}, i\omega_n) = \frac{2}{(\beta\nu)^2}\sum_{\mathbf{q}, ip_m} & \mathcal{G}^{(0)}(\mathbf{p}, ip_m)\mathcal{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip_m + i\omega_n) \\ & \times \sum_{\mathbf{k}, ik_{n'}} M_{\mathbf{k}}^2 \mathcal{D}^{(0)}(\mathbf{k}, ik_{n'})\mathcal{G}^{(0)}(\mathbf{p}+\mathbf{k}+\mathbf{q}, ip_n + i\omega_m + ik_{n'}) \\ & \times \mathcal{G}^{(0)}(\mathbf{p}+\mathbf{k}, ip_n + ik_{n'}) \end{aligned} \tag{3.215}$$

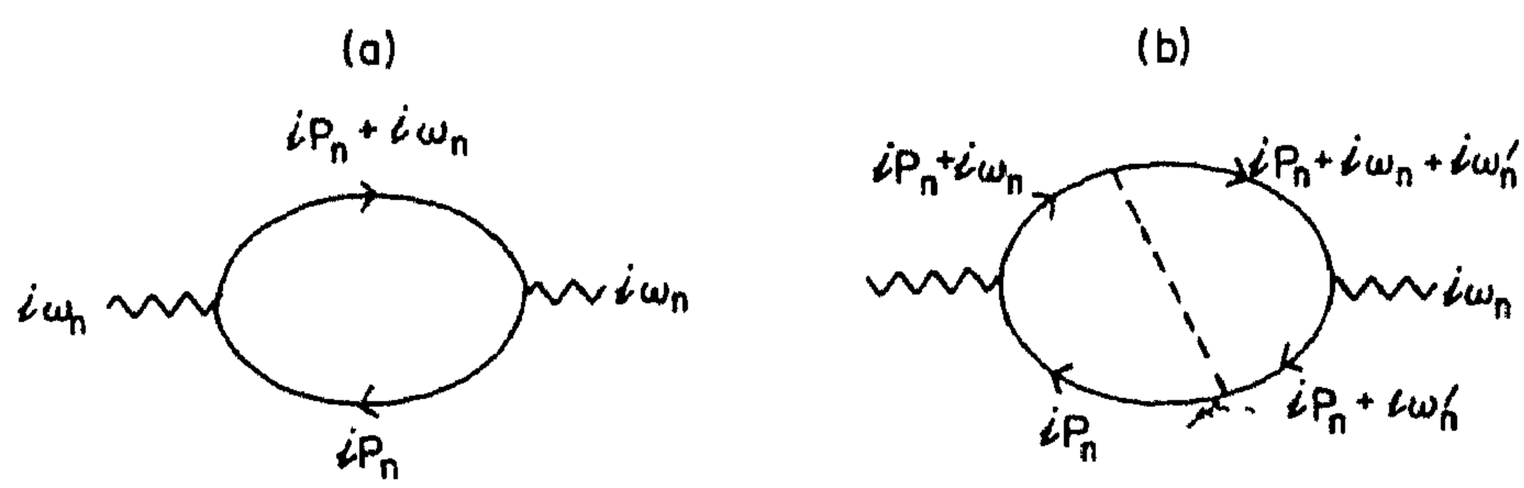


FIGURE 3.5

3.5. FREQUENCY SUMMATIONS

When using the Matsubara Green's functions, one must often evaluate frequency summations over combinations of unperturbed Green's functions. The technique for evaluating these summations is discussed for both cases of unperturbed functions and also Green's functions with self-energies. First is presented a table of results for combinations which often occur. Their derivation is later:

$$-\frac{1}{\beta}\sum_m \mathcal{D}^{(0)}(\mathbf{q}, i\omega_m)\mathcal{G}^{(0)}(\mathbf{p}, ip_n + i\omega_m) = \frac{N_\mathbf{q} + n_F(\xi_\mathbf{p})}{ip_n - \xi_\mathbf{p} + \omega_\mathbf{q}} + \frac{N_\mathbf{q} + 1 - n_F(\xi_\mathbf{p})}{ip_n - \xi_\mathbf{p} - \omega_\mathbf{q}} \tag{3.216}$$

$$\frac{1}{\beta}\sum_n \mathcal{G}^{(0)}(\mathbf{p}, ip_n)\mathcal{G}^{(0)}(\mathbf{k}, ip_n + i\omega_m) = \frac{n_F(\xi_\mathbf{p}) - n_F(\xi_\mathbf{k})}{i\omega_m + \xi_\mathbf{p} - \xi_\mathbf{k}} \tag{3.217}$$

$$-\frac{1}{\beta}\sum_n \mathcal{G}^{(0)}(\mathbf{p}, ip_n)\mathcal{G}^{(0)}(\mathbf{k}, i\omega_m - ip_n) = \frac{1 - n_F(\xi_\mathbf{p}) - n_F(\xi_\mathbf{k})}{i\omega_m - \xi_\mathbf{p} - \xi_\mathbf{k}} \tag{3.218}$$

$$\frac{1}{\beta}\sum_n \mathcal{G}^{(0)}(\mathbf{p}, ip_n) = n_F(\xi_\mathbf{p}) \tag{3.219}$$

The combination (3.216) occurs in the electron self-energy (3.212) from the electron–phonon interaction, while the combination (3.217) is found in the basic polarization diagram (3.214). The other combinations will be encountered later.

First consider the summation over a boson series, where the summation is over even integer combinations $\omega_m = 2\pi m k_B T$. For example, consider (3.216):

$$S = \frac{1}{\beta}\sum_m \frac{2\omega_\mathbf{q}}{\omega_m^2 + \omega_\mathbf{q}^2}\frac{1}{ip_n + i\omega_m - \xi_\mathbf{p}} \tag{3.220}$$

First, write this equation as

$$S = -\frac{1}{\beta}\sum_m f(i\omega_m) \tag{3.221}$$

where $f(i\omega_m)$ is the product of Green's functions in (3.220). This summation is evaluated by a contour integration. The integral has the form

$$I = \lim_{R\to\infty}\oint \frac{dz}{2\pi i} f(z) n_B(z) \tag{3.222}$$

where the contour is a large circle of radius R in the limit as $R \to \infty$. The function $n_B(z)$ is chosen to generate poles at the points $i\omega_m$ for all even integer m. The function which does this is

$$n_B(z) = \frac{1}{e^{\beta z} - 1} \tag{3.223}$$

The poles of $n_B(z)$ were discussed earlier at the beginning of the chapter, in (3.6). There are poles at the points $i2\pi m k_B T$ for all positive and negative integer m and $m = 0$. The residue at these poles is $1/\beta$. In Fig. 3.6, these poles are shown as X marks which are evenly spaced on

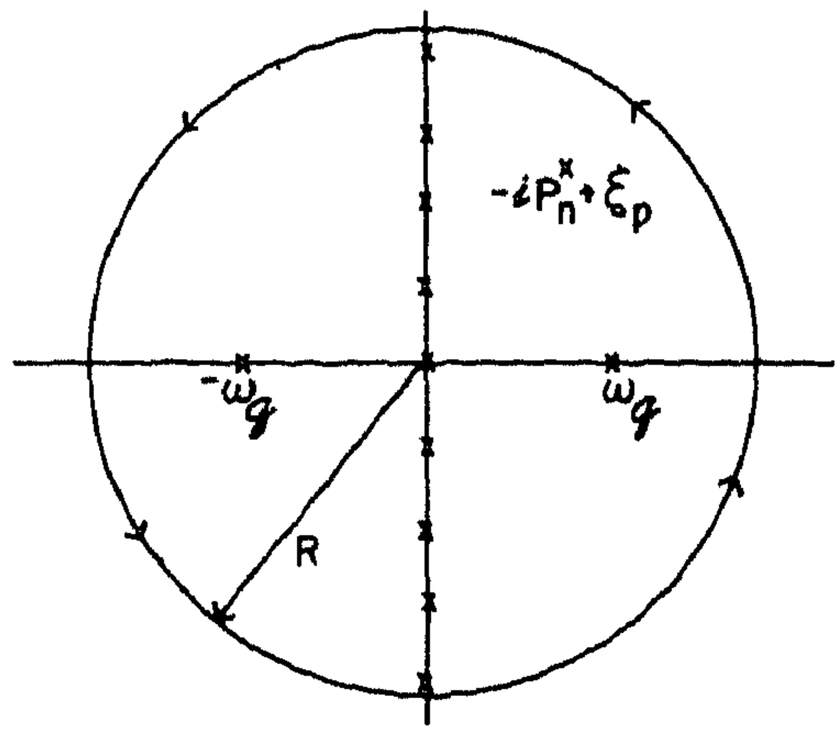


FIGURE 3.6

the vertical axis. The large circle is the contour of integration, which is a circle of radius R. The function $f(z)$ is

$$f(z) = \frac{2\omega_{\mathbf{q}}}{z^2 - \omega_{\mathbf{q}}^2} \frac{1}{ip_n + z - \xi_{\mathbf{p}}} \tag{3.224}$$

It has poles at the points $\pm\omega_{\mathbf{q}}$, which originate from the phonon Green's function. Another pole is at the point $\xi_{\mathbf{q}} - ip_n$, and this pole comes from the electron Green's function. The poles, and their residues, of the integral I in (3.222) are

$$\begin{aligned} z_m &= i2\pi m k_B T, & R_i &= \frac{1}{\beta} f(i\omega_m) \\ z_1 &= \omega_{\mathbf{q}} & R_1 &= \frac{N_{\mathbf{q}}}{ip_n - \xi_{\mathbf{p}} + \omega_{\mathbf{q}}} \\ z_2 &= -\omega_{\mathbf{q}} & R_2 &= \frac{N_{\mathbf{q}} + 1}{ip_n - \xi_{\mathbf{p}} - \omega_{\mathbf{q}}} \\ z_3 &= \xi_{\mathbf{p}} - ip_n & R_3 &= \frac{-2\omega_{\mathbf{q}} n_F(\xi_{\mathbf{p}})}{(ip_n - \xi_{\mathbf{p}})^2 - \omega_{\mathbf{q}}^2} \end{aligned} \tag{3.225}$$

where our notation is that $N_{\mathbf{q}} \equiv n_B(\omega_{\mathbf{q}})$. In the last residue, since $\exp(ip_n\beta) = -1$ then

$$n_B(\xi_{\mathbf{p}} - ip_n) = \frac{1}{e^{\beta(\xi_{\mathbf{p}} - ip_n)} - 1} = -\frac{1}{e^{\beta\xi_{\mathbf{p}}} + 1} = -n_F(\xi_{\mathbf{p}}) \tag{3.226}$$

The last residue may be rewritten in partial fractions

$$R_3 = \frac{n_F(\xi_{\mathbf{p}})}{ip_n - \xi_{\mathbf{p}} + \omega_{\mathbf{q}}} - \frac{n_F(\xi_{\mathbf{p}})}{ip_n - \xi_{\mathbf{p}} - \omega_{\mathbf{q}}} \tag{3.227}$$

The integral is evaluated by adding all these residues:

$$I = \frac{1}{\beta} \sum_m f(i\omega_m) + \frac{N_{\mathbf{q}} + n_F(\xi_{\mathbf{p}})}{ip_n - \xi_{\mathbf{p}} + \omega_{\mathbf{q}}} + \frac{N_{\mathbf{q}} + 1 - n_F(\xi_{\mathbf{p}})}{ip_n - \xi_{\mathbf{p}} - \omega_{\mathbf{q}}}$$

The integral vanishes ($I = 0$) in the limit that $R \to 0$ which gives the final result,

$$S = \frac{N_{\mathbf{q}} + n_F(\xi_{\mathbf{p}})}{ip_n - \xi_{\mathbf{p}} + \omega_{\mathbf{q}}} + \frac{N_{\mathbf{q}} + 1 - n_F(\xi_{\mathbf{p}})}{ip_n - \xi_{\mathbf{p}} - \omega_{\mathbf{q}}} \tag{3.228}$$

which is (3.216). The method of evaluating these boson series is quite simple. To evaluate a series such as (3.221) one just finds all the simple poles of $f(z)$ which are at the points z_j with residues r_j and

$$S = \sum_j r_j n_B(z_i) \tag{3.229}$$

The same procedure is used to evaluate fermion series. Then the summations are of the form

$$S = -\frac{1}{\beta}\sum_n f(ip_n) \tag{3.230}$$

where $p_n = (2n+1)\pi k_B T$ contains odd integers. One constructs the same contour integral as in (3.222). Now the function $n_F(z)$ is used since it has poles at the points ip_n with residues $-1/\beta$. The integral again is $I = 0$, so that the summation is over the residues r_i at the poles z_i of $f(z)$:

$$S = -\sum_I r_i n_F(z_i) \tag{3.231}$$

The minus sign in front occurs because the residue of the fermion $n_F(z)$ is $(-1/\beta)$, whereas it is $(1/\beta)$ for the boson $n_B(z)$.

As an example of summing a fermion series, again consider (3.216). The summation variable is changed to $p_{n'} = p_n + \omega_m$ so that the summation is now

$$S = -\frac{1}{\beta}\sum_{n'} \mathcal{D}^{(0)}(\mathbf{q}, ip_{n'} - ip_n)\mathcal{G}^{(0)}(\mathbf{p}, ip_{n'}) \tag{3.232}$$

$$= \frac{1}{\beta}\sum_{n'} \frac{2\omega_{\mathbf{q}}}{(p_{n'} - p_n)^2 + \omega_{\mathbf{q}}^2}\frac{1}{ip_{n'} - \xi_{\mathbf{p}}} \tag{3.233}$$

$$f(z) = \frac{2\omega_{\mathbf{q}}}{(z - ip_n)^2 - \omega_{\mathbf{q}}^2}\frac{1}{z - \xi_{\mathbf{p}}} \tag{3.234}$$

$p_{n'}$ contains odd integers, since it is a fermion frequency. The function $f(z)$ has the following poles and residues:

$$\begin{aligned} z_1 &= \xi_{\mathbf{p}}, & R_1 &= \frac{2\omega_{\mathbf{q}} n_F(\xi_{\mathbf{p}})}{(\xi_{\mathbf{p}} - ip_n)^2 - \omega_{\mathbf{q}}^2} \\ z_2 &= ip_n - \omega_{\mathbf{q}}, & R_2 &= \frac{-n_F(ip_n - \omega_{\mathbf{q}})}{ip_n - \xi_{\mathbf{p}} - \omega_{\mathbf{q}}} \\ z_3 &= ip_n + \omega_{\mathbf{q}}, & R_3 &= \frac{+n_F(ip_n + \omega_{\mathbf{q}})}{ip_n - \xi_{\mathbf{p}} + \omega_{\mathbf{q}}} \end{aligned} \tag{3.235}$$

The last two thermal factors may be simplified:

$$n_F(ip_n - \omega_{\mathbf{q}}) = \frac{1}{e^{\beta(ip_n - \omega_{\mathbf{q}})} + 1} = \frac{1}{1 - e^{-\beta\omega_{\mathbf{q}}}} = N_{\mathbf{q}} + 1 \tag{3.236}$$

$$n_F(ip_n + \omega_{\mathbf{q}}) = -N_{\mathbf{q}} \tag{3.237}$$

The final result is obtained, as in (3.231), by adding these three residues, which gives again the same answer (3.228).

Three summations were listed at the beginning of the section. The first has been derived twice. The next two, (3.217) and (3.218), are left as an exercise for the student. The fourth result (3.219) is now explained

$$\frac{1}{\beta}\sum_n \mathcal{G}^{(0)}(\mathbf{p}, ip_n) = n_F(\xi_{\mathbf{p}})$$

The left-hand side of this equation is most familiar as the Fourier transform of the Matsubara Green's function:

$$\mathcal{G}^{(0)}(\mathbf{p}, \tau) = \frac{1}{\beta}\sum_n e^{-ip_n\tau}\mathcal{G}^{(0)}(\mathbf{p}, ip_n) = -{}_0\langle T_\tau C_{\mathbf{p}\sigma}(\tau)C^\dagger_{\mathbf{p}\sigma}(0)\rangle$$

The result (3.219) is just the limit as $\tau \to 0$. This limit is ambiguous, since a different result is obtained if $\tau = 0$ is approached from the positive or negative direction:

$$\mathcal{G}^{(0)}(\mathbf{p}, \tau = 0^+) = -{}_0\langle C_{\mathbf{p}\sigma}C^\dagger_{\mathbf{p}\sigma}\rangle = -[1 - n_F(\xi_{\mathbf{p}})] \tag{3.238}$$

$$\mathcal{G}^{(0)}(\mathbf{p}, \tau = 0^-) = {}_0\langle C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}\rangle = n_F(\xi_{\mathbf{p}}) \tag{3.239}$$

The result (3.219) is merely the convention of adopting the limit $\tau = 0^-$. This choice has been discussed several times before. The same result at zero temperature was expressed in (2.86). It is the convention that two operators at equal time are taken in the order which gives the number operator:

$$-{}_0\langle T_\tau C_{\mathbf{p}\sigma}(\tau)C^\dagger_{\mathbf{p}\sigma}(\tau)\rangle = {}_0\langle C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}\rangle = n_F(\xi_{\mathbf{p}}) \tag{3.240}$$

The Matsubara sum in the preceding example was very easy. Summations which contain only $\mathcal{G}^{(0)}$ and $\mathcal{D}^{(0)}$ always cause the contour integral I to have just poles. If the summation contained $\mathcal{G}$, instead of $\mathcal{G}^{(0)}$, then the I integral has branch cuts. Next is shown how to treat this more difficult case.

Consider the example of evaluating the summation (3.216) except that the Green's function for the electron contains some self-energy terms:

$$\mathcal{G}(\mathbf{p}, ip_n) = \frac{1}{ip_n - \xi_{\mathbf{p}} - \Sigma(\mathbf{p}, ip_n)} \tag{3.241}$$

Letting $ip_n = z$ then the above function has branch cuts along the real z axis. These branch cuts arise because of the self-energies $\Sigma(\mathbf{p}, z)$, as discussed in Sec. 3.3. The contour integrals used to evaluate the frequency summations become more complicated because of the branch cuts.

The most general possibility when evaluating a summation over Matsubara frequencies is to have all the Green's functions fully dressed. Then it is easiest to proceed by first expressing all the Green's functions in the Lehmann representation. For example, evaluate

$$S = -\frac{1}{\beta}\sum_m \mathcal{D}(\mathbf{q}, i\omega_m)\mathcal{G}(\mathbf{p}, ip_n + i\omega_m) \tag{3.242}$$

Express each Green's function as a frequency integral over its respective spectral functions:

$$\mathcal{D}(\mathbf{q}, i\omega_m) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\frac{B(\mathbf{q}, \omega)}{i\omega_m - \omega} \tag{3.243}$$

$$\mathcal{G}(\mathbf{p} + \mathbf{q}, ip_n + i\omega_m) = \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi}\frac{A(\mathbf{p} + \mathbf{q}, \varepsilon)}{ip_n + i\omega_m - \varepsilon} \tag{3.244}$$

and the summation S becomes

$$S = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} B(\mathbf{q}, \omega) \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} A(\mathbf{p}+\mathbf{q}, \varepsilon) S_0(ip_n, \omega, \varepsilon)$$

$$S_0(ip_n, \omega, \varepsilon) = -\frac{1}{\beta} \sum_m \frac{1}{i\omega_m - \omega} \frac{1}{ip_n + i\omega_m - \varepsilon} = \frac{n_B(\omega) + n_F(\varepsilon)}{ip_n + \omega - \varepsilon}$$

$$S = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} B(\mathbf{q}, \omega) \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} A(\mathbf{p}+\mathbf{q}, \varepsilon) \frac{n_B(\omega) + n_F(\varepsilon)}{ip_n + \omega - \varepsilon}$$

The summation $S_0(ip_n, \omega, \varepsilon)$ is now the easy type which is evaluated using noninteracting Green's functions. The final result provides for the most general case of fully interacting Green's functions. In general, any frequency summation can be done by expressing all the Green's functions in their Lehmann representation, which results in a summation of the easy type listed in (3.216) to (3.219).

The result (3.216) has now been derived in this section by several different methods. Now we explain the physics behind this result. It is a bit subtle. The virtue of the Matsubara method is that these subtleties are automatically handled correctly. It is a machinery which cranks out the correct thermal occupation factors n_F and $N_{\mathbf{q}}$ even in complicated cases.

The summation (3.216) arises from the electron–phonon interaction and is a phonon contribution to the self-energy of the electron. From the rules for constructing diagrams, the one-phonon self-energy of the electron is

$$\Sigma^{(1)}(\mathbf{p}, ip_n) = \frac{1}{\nu} \sum_{\mathbf{q}} M_{\mathbf{q}}^2 \left[\frac{N_{\mathbf{q}} + 1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})}{ip_n - \xi_{\mathbf{p}+\mathbf{q}} - \omega_{\mathbf{q}}} + \frac{N_{\mathbf{q}} + n_F(\xi_{\mathbf{p}+\mathbf{q}})}{ip_n - \xi_{\mathbf{p}+\mathbf{q}} + \omega_{\mathbf{q}}} \right] \tag{3.245}$$

This result is similar to what is obtained from second-order perturbation theory. The standard quantum mechanical expression for this quantity is

$$\Delta E^{(2)} = \sum_I \frac{|\langle I | H_{ep} | i \rangle|^2}{E_i - E_I} \tag{3.246}$$

where $|i\rangle$ is the initial state of the system, and the summation $|I\rangle$ is over possible intermediate states. It is assumed that the initial state is an electron in state $\mathbf{p}$ with energy $\varepsilon_{\mathbf{p}}$. The electron–phonon interaction has two terms. One of these is $M_{\mathbf{q}} C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} a^{\dagger}_{-\mathbf{q}}$, which describes a process where an electron scatters from $\mathbf{p}$ to $\mathbf{p}+\mathbf{q}$ while creating a phonon of momentum $-\mathbf{q}$. This contribution from second-order perturbation theory is

$$\frac{1}{\nu} \sum_{\mathbf{q}} M_{\mathbf{q}}^2 \frac{(N_{\mathbf{q}} + 1)[1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})]}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} - \omega_{\mathbf{q}}} \tag{3.247}$$

The factor $(N_{\mathbf{q}} + 1)$ comes from the phonon creation operator, $a^{\dagger}_{-\mathbf{q}} |n_{-\mathbf{q}}\rangle = \sqrt{n_{-\mathbf{q}} + 1} |n_{-\mathbf{q}} + 1\rangle$. The thermal average is used for $n_{-\mathbf{q}}$, which is $N_{\mathbf{q}}$. Similarly, the factor $[1 - n_F]$ is the probability that the electron state $\mathbf{p}+\mathbf{q}$ is empty, so that the operator $C^{\dagger}_{\mathbf{p}+\mathbf{q}\sigma}$ can create an electron in that state. There is also a factor $n_F(\xi_{\mathbf{p}})$, which is the probability that $\mathbf{p}$ is occupied with an electron. It is assumed this factor is unity if we are trying to calculate the properties of an electron in that state. The energy denominator gives the difference between the initial energy $\varepsilon_{\mathbf{p}}$ and the energy in the intermediate state $\varepsilon_{\mathbf{p}+\mathbf{q}} + \omega_{\mathbf{q}}$ which has an additional phonon.

Similarly, the interaction term which destroys a phonon is $M_{\mathbf{q}} C^{\dagger}_{\mathbf{p}+\mathbf{q}\sigma} C_{\mathbf{p}\sigma} a_{\mathbf{q}}$. In second order it contributes an energy term

$$\frac{1}{\nu}\sum_{\mathbf{q}} M_{\mathbf{q}}^2 \frac{N_{\mathbf{q}}[1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})]}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega_{\mathbf{q}}} \tag{3.248}$$

Here the factor $N_{\mathbf{q}}$ is the matrix element of the phonon destruction operator $a_{\mathbf{q}}|n_{\mathbf{q}}\rangle = \sqrt{n_{\mathbf{q}}}|n_{\mathbf{q}} - 1\rangle$. The electron occupation factor $[1 - n_F]$ is the same as before. The phonon energy in the denominator has changed sign, because there is one less phonon in the intermediate state.

These are the only two terms in the Hamiltonian. The summation of these two contributions, however, does not equal (3.245). In fact, some important processes have been omitted. There are two more terms. They arise from the other electrons in the system with the same spin state σ. These other electrons are also trying to emit and absorb phonons, to alter their energy. The other electrons may not, in this process, use the state $\mathbf{p}$ as their intermediate state. For example, if another electron starts in $\mathbf{p}'$ and scatters to $\mathbf{p}' + \mathbf{q}$, we may not have $\mathbf{p} = \mathbf{p}' + \mathbf{q}$. They may not use $\mathbf{p}$ because our electron, whose energy we are trying to calculate, is occupying it already. The other electrons of the system have a reduced second-order self-energy because our electron is in the state $\mathbf{p}$. This energy reduction is associated with $\mathbf{p}$. It is assigned to the self-energy of the electron in $\mathbf{p}$, because the system would not have it if there were no electron in $\mathbf{p}$. A calculation of the self-energy of an electron in the state $(\mathbf{p}, \sigma)$ must calculate how much energy is deprived to the other electrons. Assume they start in $\mathbf{p} + \mathbf{q}$ initially and get scattered into $\mathbf{p}$. In analogy with the two terms above, the self-energy of a particle in the state $\mathbf{p} + \mathbf{q}$, for emitting or absorbing a phonon and going into $\mathbf{p}$, is

$$\frac{1}{\nu}\sum_{\mathbf{q}} M_{\mathbf{q}}^2 [1 - n_F(\xi_{\mathbf{p}})] n_F(\xi_{\mathbf{p}+\mathbf{q}}) \left(\frac{N_{\mathbf{q}} + 1}{\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}} - \omega_{\mathbf{q}}} + \frac{N_{\mathbf{q}}}{\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}} + \omega_{\mathbf{q}}} \right) \tag{3.249}$$

The factor $[1 - n_F]$ is the probability that $\mathbf{p}$ is empty. This factor is dropped, since $\mathbf{p}$ is occupied. The factor $n_F(\xi_{\mathbf{p}+\mathbf{q}})$ is the probability that there is an electron in $\mathbf{p} + \mathbf{q}$; if there is none, then the electron in $\mathbf{p}$ is not depriving any electron of energy. The calculation assumes that the energy does not depend on spin, so the occupation factors n_F should not depend on spin either. These last two terms are then subtracted from the first two to give the result, from second-order perturbation theory,

$$\Delta E^{(2)} = \frac{1}{\nu}\sum_{\mathbf{q}} M_{\mathbf{q}}^2 \left[\frac{N_{\mathbf{q}} + 1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} - \omega_{\mathbf{q}}} + \frac{N_{\mathbf{q}} + n_F(\xi_{\mathbf{p}+\mathbf{q}})}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega_{\mathbf{q}}} \right] \tag{3.250}$$

The Green's function theory (3.245) changes this result only a little bit. It allows that the original electron may have had wave vector $\mathbf{p}$, but its "energy" is $ip_n + \mu$ and not $\varepsilon_{\mathbf{p}}$. It keeps the energy and momentum as separate variables, at least at this stage in the calculation. An obvious improvement in the calculation is to make the intermediate state energy not $\varepsilon_{\mathbf{p}+\mathbf{q}}$ but some other energy variable which reflects the possible effects of interaction. The difficulty with such a procedure is that changes in $\varepsilon_{\mathbf{p}+\mathbf{q}}$ also require changes in the occupation probability $n_F(\xi_{\mathbf{p}+\mathbf{q}})$. Later it will be shown that one subset of diagrams, when summed to infinity, yield exactly this approximation; the intermediate state energies are treated exactly also.

In many-particle systems, the interactions which alter energy states of particles usually lead to subtle counting problems about where energy has been added or subtracted from the

system. The Matsubara method has the virtue that it cranks out the correct answers automatically.

3.6. LINKED CLUSTER EXPANSIONS

Another method for evaluating correlation functions is called the *linked cluster method* or the *cumulant expansion* (Abrikosov *et al.*, 1963; Brout and Carruthers, 1963). It is the method used to evaluate the thermodynamic potential Ω. However, the method has also been applied to evaluating Green's functions of time $G(\mathbf{p}, t)$ for a few problems, which represents an alternative to the usual Dyson equation approach. Both methods—Dyson's equation and linked cluster—would give the same result if all terms were summed exactly. However, in the usual case, when only a few terms in a perturbation expansion are evaluated and the rest are neglected, Dyson's equation and the linked cluster method are both approximate solutions. In some special cases the linked cluster expansion may be the better approximation. The art of choosing which method to apply to a given problem is a matter of judgment, which comes from experience.

3.6.1. Thermodynamic Potential

The thermodynamic potential is found from the equation

$$e^{-\beta\Omega} = \mathrm{Tr}(e^{-\beta K}) = \mathrm{Tr}[e^{-\beta K_0} S(\beta)] \tag{3.251}$$

This quantity is useful to evaluate by itself. After it is found, perhaps only approximately, some thermodynamic results are determined by taking its derivatives; for example, if N_e is the number of electrons

$$-\frac{\partial\Omega}{\partial\mu} = \langle N_e \rangle = \bar{N}_e \tag{3.252}$$

$$\frac{\partial(\beta\Omega)}{\partial\beta} = \langle K \rangle = U - \mu\bar{N}_e = \Omega + TS \tag{3.253}$$

$$m\frac{\partial\Omega}{\partial m} = \left\langle \sum_i \frac{p_i^2}{2m} \right\rangle = \sum_{\mathbf{p}} n_{\mathbf{p}} \frac{p^2}{2m} \tag{3.254}$$

In the last identity (3.254), the function $n_{\mathbf{p}}$ is meant to be the number of electrons with momentum $\mathbf{p}$. This quantity was defined earlier, in (3.135), for an interacting system.

Usually there is a Hamiltonian which is written as

$$H = H_0 + V \tag{3.255}$$

$$K = K_0 + V = (H_0 - \mu N_e) + V \tag{3.256}$$

where H_0 is exactly solvable. In the interaction representation, the thermodynamic potential is given by the right-hand side of (3.251). If there are no interactions and $V = 0$, then the definition of Ω_0 is

$$e^{-\beta\Omega_0} = \mathrm{Tr}(e^{-\beta K_0}) \tag{3.257}$$

For a system of electrons and phonons,

$$K_0 = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \sum_{\mathbf{q}} \omega_{\mathbf{q}} (a^\dagger_{\mathbf{q}} a_{\mathbf{q}} + \tfrac{1}{2}) \tag{3.258}$$

the trace is simple since the noninteracting eigenstates can be used to expand the trace eigenstates:

$$e^{-\beta\Omega_0} = \Pi_{\mathbf{p}}(1 + e^{-\beta\xi_{\mathbf{p}}})^2 \Pi_{\mathbf{q}} \left(\frac{e^{-(1/2)\beta\omega_{\mathbf{q}}}}{1 - e^{-\beta\omega_{\mathbf{q}}}} \right) \tag{3.259}$$

Take the logarithm of both sides of this equation,

$$\beta\Omega_0 = -2 \sum_{\mathbf{p}} \ln(1 + e^{-\beta\xi_{\mathbf{p}}}) + \sum_{\mathbf{q}} [\ln(1 - e^{-\beta\omega_{\mathbf{q}}}) + \tfrac{1}{2}\beta\omega_{\mathbf{q}}] \tag{3.260}$$

The factor of 2 in front of the electron term is the spin degeneracy. The last term on the right is the zero-point energy of the phonons. Change the summations to integrals, say for the phonons

$$\sum_{\mathbf{q}} F(\mathbf{q}) = \mathrm{v} \int \frac{d^3q}{(2\pi)^3} F(\mathbf{q}) \tag{3.261}$$

and a factor appears of the volume v. The thermodynamic potential is proportional to the volume of the system. This result is also apparent from Eqns. (3.252) to (3.254), since the right-hand side in each case is proportional to the volume or to the number of particles.

The goal is to evaluate (3.251) including $S(\beta)$. The interaction V will add some correction terms to Ω_0 which changes it into Ω. The method of evaluating this correlation function is by the S-matrix expansion given earlier in (3.195):

$$e^{-\beta\Omega} = e^{-\beta\Omega_0} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_n \, {}_0\langle T_\tau \hat{V}(\tau_1) \cdots \hat{V}(\tau_n) \rangle$$

In evaluating the right-hand side, all diagrams are included, whether connected or disconnected, since there is no other series at hand to cancel the disconnected diagrams. The disconnected diagrams are necessary for the resummation procedure. Basically, the terms in the S-matrix series are examined in order to find some convenient method of resumming them. It is efficient to evaluate each different type of diagram only once. The way of doing this expansion is the linked cluster, or cumulant, expansion.

The answer is stated first. The following discussion will try to make the answer appear plausible. First, introduce the parameter λ. It multiplies the potential V, so that everywhere there was a factor of V, or $V(\tau)$, now write λV or $\lambda V(\tau)$. The parameter λ is actually just unity, but it is used to keep track of the number of times the potential occurs in each term in the S matrix. The S-matrix expansion in (3.195) may be written as

$$e^{-\beta\Omega} = e^{-\beta\Omega_0} \sum_{n=0}^{\infty} \lambda^n W_n \tag{3.262}$$

$$W_n = \frac{(-1)^n}{n!} \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_n \, {}_0\langle T_\tau \hat{V}(\tau_1) \cdots \hat{V}(\tau_n) \rangle \tag{3.263}$$

The basic linked cluster theorem is that this series can be resummed into

$$e^{-\beta\Omega} = \exp\left(-\beta\Omega_0 + \sum_{l=1}^{\infty} \lambda^l U_l\right) \tag{3.264}$$

$$U_l = \frac{(-1)^l}{l} \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_l \, {}_0\langle T_\tau \hat{V}(\tau_1) \cdots \hat{V}(\tau_l)\rangle \tag{3.265}$$

where U_l contains just different, connected diagrams: The thermodynamic potential is obtained by setting $\lambda = 1$ in (3.264)

$$\Omega = \Omega_0 - \frac{1}{\beta} \sum_{l=1}^{\infty} U_l \tag{3.266}$$

The basic theorem is that the thermodynamic potential is the summation of the different, connected diagrams. The factor $1/l$ in (3.265) turns out to be quite a nuisance, as is discussed later. It makes further summation difficult, except by a trick.

The theorem will be proved below. First it is helpful to give a simple example. Consider the perturbation V to be the electron–phonon interaction:

$$V = \sum_{\mathbf{q}\mathbf{p}\sigma} \frac{M_{\mathbf{q}}}{\sqrt{\mathcal{V}}} A_{\mathbf{q}} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{3.267}$$

For this example, only terms with even n occur in the series, since the expectation value of $\langle A^n_{\mathbf{q}}\rangle$ is zero if n is odd. The first nonzero term in the series, after $n = 0$, is $n = 2$:

$$W_2 = U_2 = \frac{1}{2} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \, {}_0\langle T_\tau \hat{V}(\tau_1)\hat{V}(\tau_2)\rangle \tag{3.268}$$

This disconnected diagram is not evaluated here. Its Feynman diagram is shown in Fig. 3.7(a). There are two electron lines, and the two vertices are connected by a phonon line. Next consider the term with $n = 4$:

$$W_4 = \frac{1}{4!} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \int_0^\beta d\tau_3 \int_0^\beta d\tau_4 \, {}_0\langle T_\tau \hat{V}(\tau_1)\hat{V}(\tau_2)\hat{V}(\tau_3)\hat{V}(\tau_4)\rangle$$

All contributions for $n = 4$ consist of diagrams with four electron lines and two phonon lines. Some are shown in Fig. 3.7(d). However, there also occurs the diagrams shown in Fig. 3.7(b) where the single bubble occurs twice. In fact, there are three identical terms in W_4 which have a pair of disconnected bubbles. The factor of 3 occurs because, if one chooses τ_1 as one of the vertices of a bubble, the other τ variable for the same bubble is either τ_2, τ_3, or τ_4, and these three choices each give the same contribution. This factor is

$$\frac{3}{24} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \, {}_0\langle T_\tau \hat{V}(\tau_1)\hat{V}(\tau_2)\rangle \int_0^\beta d\tau_3 \int_0^\beta d\tau_4 \, {}_0\langle T_\tau \hat{V}(\tau_3)\hat{V}(\tau_4)\rangle = \tfrac{1}{2} U_2^2$$

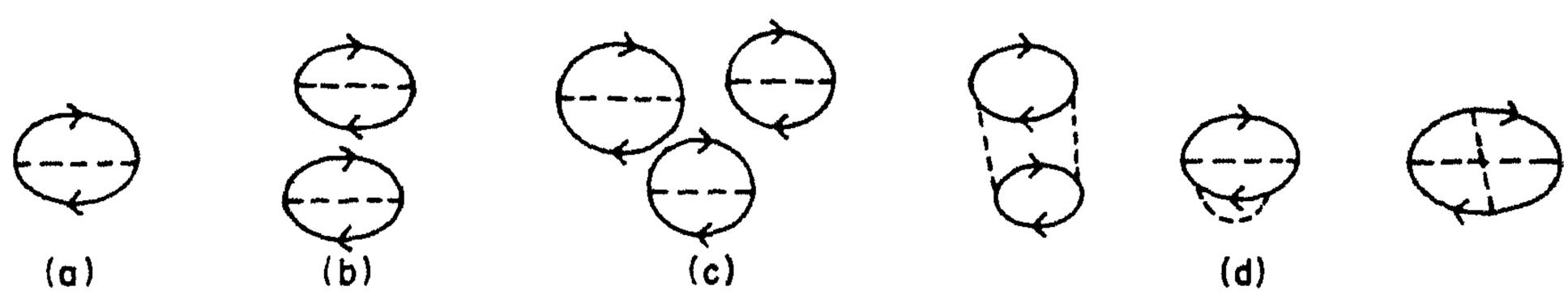


FIGURE 3.7

Similarly, it can be shown that each of the connected diagrams in Fig. 3.7(d) occurs six times, where the factor of 6 comes from different ways to shift the dummy variables around. The final result is

$$W_4 = \tfrac{1}{2}U_2^2 + U_4 \tag{3.269}$$

Next consider terms with $n = 6$. There will be many connected diagrams which contain six electron lines and three phonon lines; try to draw them all. They give U_6. There are also disconnected diagrams. There will be diagrams containing three bubbles, as shown in Fig. 3.7(c)—in fact there will be 15 of these, all equal to each other. There will also be disconnected diagrams which contain one four-line term from Fig. 3.7(d) and one from Fig. 3.7(a). The total is

$$W_6 = \frac{1}{3!}U_2 + U_2U_4 + U_6 \tag{3.270}$$

The terms in U_2 appear to be generating the series

$$1 + U_2 + \frac{1}{2!}U_2^2 + \frac{1}{3!}U_2^3 + \cdots = e^{U_2} \tag{3.271}$$

There is a simple argument which shows that the nth term is really appropriate for the exponential series. Since n is even, let $n = 2m$, and there are m bubbles of the type shown in Fig. 3.7(a). How many different ways are there to get m bubbles? Take the first variable, τ_1. It may be paired in the bubble,

$$\begin{aligned} U_2 = \frac{1}{2}\sum_{\mathbf{q}\mathbf{p}\sigma} M_{\mathbf{q}}^2 \int_0^\beta d\tau_1 \int_0^\beta d\tau_j \mathcal{D}^{(0)}(\mathbf{q}, \tau_1 - \tau_j) \\ \times \mathcal{G}^{(0)}(\mathbf{p}, \tau_1 - \tau_j)\mathcal{G}^{(0)}(\mathbf{p}+\mathbf{q}, \tau_j - \tau_1) \end{aligned} \tag{3.272}$$

with the variable τ_j which is any of the other $2m - 1$ variables. Next, take any of the $2m - 2$ variables left after pairing τ_1 and τ_j, say τ_s, and it may be paired with any of the $2m - 3$ left. The number of different arrangements is

$$(2m-1)(2m-3)(2m-5)\cdots 5\cdot 3\cdot 1 = \frac{(2m)!}{2^m m!} \tag{3.273}$$

This factor in W_n is

$$W_n = \frac{(2m)!}{2^m m!}\frac{1}{(2m)!}\left[\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \, {}_0\langle T_\tau \hat{V}(\tau_1)\hat{V}(\tau_2)\rangle\right]^m = \frac{1}{m!}(U_2)^m$$

which is the form proposed in (3.271). The term U_2 has been shown to sum correctly. It must be shown that the other contributions also sum correctly. This proof is not too hard. It is more subtle to show that all the cross products such as U_2U_4 occur with just the right counting.

Another identity is:

$$U_l = \frac{(-1)^l}{l!}\int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_l \, {}_0\langle T_\tau \hat{V}(\tau_1)\cdots\hat{V}(\tau_l)\rangle_{\text{connected}} \tag{3.274}$$

The same U_l is in (3.265). Comparing these two expressions shows that for each term l there are $(l-1)!$ contributions alike, so that the prefactor in front of the number of different connected diagrams is $(l-1)!/l! = 1/l$. The proof is just an exercise in combinatorial

counting. The S matrix is expanded in terms of Green's functions which yields expressions such as

$$U_l = \frac{1}{l!}\int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_l \mathcal{G}^{(0)}(\mathbf{p}_1, \tau_1 - \tau_j)\mathcal{G}^{(0)}(\mathbf{p}_2, \tau_2 - \tau_k)\cdots \tag{3.275}$$

The combinatorial question is: how many different combinations are really alike if the dummy variables τ_1, τ_j, τ_2, etc., are just relabeled? The answer is $(l-1)!$. The reader should try to work some examples to convince oneself that it works. The trivial case is $l = 2$ so that $(l-1)! = 1$ was done in (3.272). A more complex example is shown in Fig. 3.8(a), where the $(4-1)! = 6$ different terms are shown for one of the different diagrams with $l = 4$. The numbers 1, 2, 3, 4 refer to the τ variables: $\tau_1, \tau_2, \tau_3, \tau_4$. These six are the only different arrangements. One might ask why the figures are drawn such that τ_1 is always the upper left-hand corner. Why not the labeling shown in Fig. 3.8(b)? The answer is that this is exactly the same term as the first one in Fig. 3.8(a). They are topologically identical but just drawn upside down and backwards, or rotated by 180°. However, and most importantly, both diagrams yield identical results when written down with Green's functions, which is the ultimate test of whether they are topologically identical.

In the prior section, the Green's function $\mathcal{G}(\mathbf{p}, \tau)$ was evaluated by a similar expansion technique. There the connected diagrams obeyed the counting rules that

$$\begin{aligned}\frac{1}{l!}\ {}_0\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau)\hat{V}(\tau_1)\cdots\hat{V}(\tau_l)\hat{C}^\dagger_{\mathbf{p}\sigma}(0)\rangle_{\text{connected}} \\ = {}_0\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau)\hat{V}(\tau_1)\cdots\hat{V}(\tau_l)\hat{C}^\dagger_{\mathbf{p}\sigma}(0)\rangle_{\text{different connected}}\end{aligned} \tag{3.276}$$

Here there are $l!$ arrangements which are identical. In the linked cluster method there are only $(l-1)!$. The difference is in $\mathcal{G}(\mathbf{p}, \tau)$ there is a reference variable τ, so that arrangements are considered with respect to this variable. There turn out to be more of them by a factor of l.

Next consider the evaluation of the thermodynamic potential. It is obtained by evaluating the series of terms in (3.266). An example is to evaluate the series of diagrams shown in Fig. 3.9 for the electron–phonon interaction. They can be summed to give a simple answer. Of course, this series is not the exact answer, since there are additional terms in which the bubbles may also have internal phonon lines. The rules for constructing diagrams for the terms in the thermodynamic potential are similar to those in Sec. 3.5. One follows all those

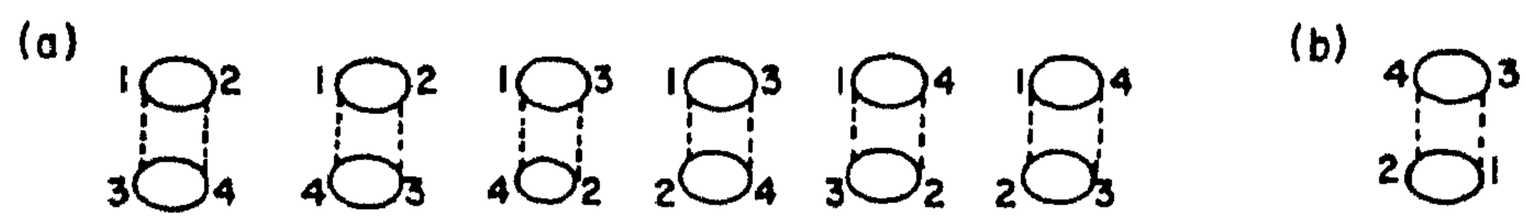


FIGURE 3.8

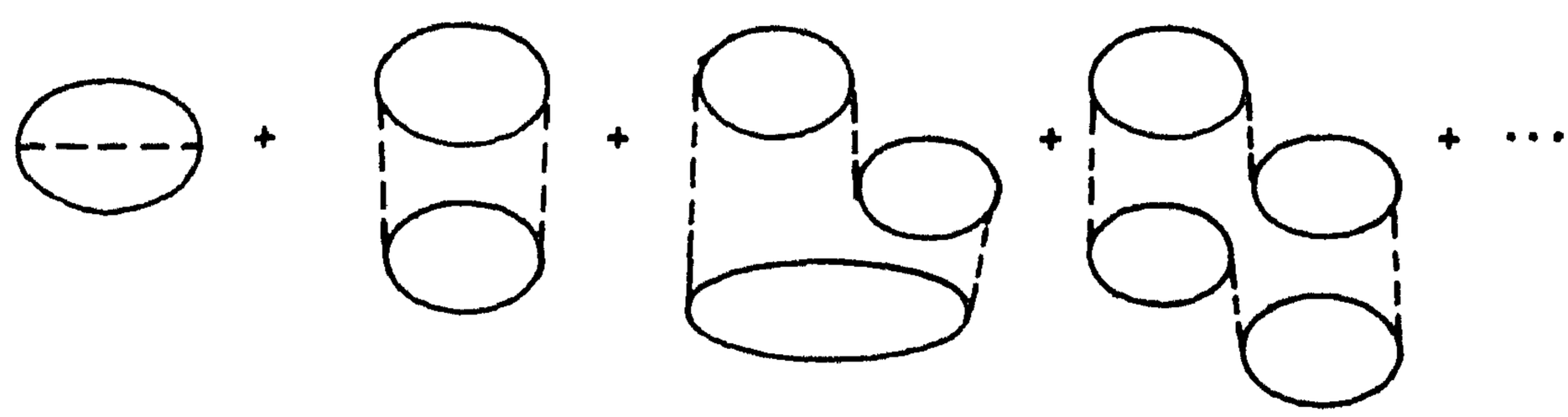


FIGURE 3.9

rules and then multiplies the result by β/l. The $1/l$ factor is just the one which occurs in (3.265), which has been explained already. The factor β comes from the extra τ integration. For example, in (3.272), which is the first bubble, the argument is only a function of $\tau_1 - \tau_j$:

$$\int_0^\beta d\tau_1 \int_0^\beta d\tau_j f(\tau_1 - \tau_j) \tag{3.277}$$

Using the Fourier expansion

$$f(\tau_1 - \tau_j) = \frac{1}{\beta}\sum_n e^{-iq_n(\tau_1-\tau_j)} f(iq_n) \tag{3.278}$$

the first integration gives

$$\frac{1}{\beta}\int_0^\beta d\tau_j e^{iq_n\tau_j} = \delta_{n=0} \tag{3.279}$$

while the second just gives β. Each term has this extra τ integration, which gives the one extra factor of β. For the first bubble in Fig. 3.9

$$U_2 = \frac{\beta}{2}\frac{\lambda^2}{\beta}\sum_{\mathbf{q},iq_n} M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n) P^{(1)}(\mathbf{q}, iq_n) \tag{3.280}$$

where the polarization diagram $P^{(1)}(\mathbf{q}, iq_n)$ is for the single bubble, which was already evaluated in (3.214):

$$P^{(1)}(\mathbf{q}, iq_n) = \frac{2}{\beta\nu}\sum_{\mathbf{p},ip_m} \mathcal{G}^{(0)}(\mathbf{p}, ip_m)\mathcal{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip_m + iq_n) \tag{3.281}$$

The second term in Fig. 3.9 has two bubbles connected by two phonon lines. Using the rules for constructing diagrams gives

$$U_4 = \frac{\beta}{4}\frac{\lambda^4}{\beta}\sum_{\mathbf{q},iq_n} [M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n) P^{(1)}(\mathbf{q}, iq_n)]^2 \tag{3.282}$$

Momentum and frequency conservation requires both phonons to have the same variables $(\mathbf{q}, iq_n)$, and both polarization diagrams are also just functions of this combination. The term with n bubbles and n phonon lines is

$$U_{2n} = \frac{\beta}{2n}\frac{\lambda^{2n}}{\beta}\sum_{\mathbf{q},iq_n} [M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n) P^{(1)}(\mathbf{q}, iq_n)]^n \tag{3.283}$$

It is simple to sum the series:

$$\Omega - \Omega_0 = -\frac{1}{\beta}\sum_{n=1}^{\infty} U_{2n} = \frac{1}{2\beta}\sum_{\mathbf{q},iq_n} \ln[1 - \lambda^2 M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n) P^{(1)}(\mathbf{q}, iq_n)] \tag{3.284}$$

The right-hand side of (3.284) is a correction to the thermodynamic potential. It is not exact, as explained above, since only some of the diagrams are included. The right-hand side is proportional to the volume ν, as is obvious when the summation over $\mathbf{q}$ is changed to an integration as in (3.261). Similarly, the summation over $\mathbf{p}$ in $P^{(1)}(\mathbf{q}, iq_n)$ can be changed to an integration:

$$\frac{1}{\nu}\sum_{\mathbf{p}} = \int \frac{d^3p}{(2\pi)^3} \tag{3.285}$$

The argument of the logarithm is not dependent on ν, and ν enters only by multiplying the result.

The answer is a logarithm. The summation over iq_n needs to be evaluated. This summation is inconvenient (but still possible) when the important functions are inside the argument of the logarithm. There is a standard trick for eliminating the logarithm, although it does not eliminate it but just disguises it. Treat $\eta = \lambda^2$ as a variable of integration,

$$\int_0^{\lambda^2} d\eta \frac{\Lambda}{1-\eta\Lambda} = -\ln[1-\lambda^2\Lambda] \tag{3.286}$$

$$\Lambda = M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n) P^{(1)}(\mathbf{q}, iq_n) \tag{3.287}$$

Then for $\lambda = 1$

$$\Omega - \Omega_0 = -\frac{\nu}{2\beta} \sum_{iq_n} \int \frac{d^3q}{(2\pi)^3} \int_0^1 d\eta \frac{M_{\mathbf{q}}^2 \mathcal{D}^{(0)} P^{(1)}}{1-\eta M_{\mathbf{q}}^2 \mathcal{D}^{(0)} P^{(1)}} \tag{3.288}$$

The logarithm has been eliminated, but another problem is introduced. The right-hand side must be evaluated for each value of η and then the integral is taken over η. Since η enters in the same way as a coupling constant, this step is called a *coupling constant integration*. The integration can be taken outside of the summation over iq_n, which makes this summation easier in some cases.

There are some other manipulations on this formula, which put it into a form which appears simpler. First, define the phonon self-energy as

$$\pi^{(1)}(\eta, \mathbf{q}, iq_n) = \eta M_{\mathbf{q}}^2 P^{(1)}(\mathbf{q}, iq_n) \tag{3.289}$$

where the coupling constant η is included in the definition. The superscript (1) indicates that this expression is an approximate self-energy, which only includes the single-electron bubble. The phonon Green's function is

$$\mathcal{D}^{(1)}(\eta, \mathbf{q}, iq_n) = \frac{\mathcal{D}^{(0)}}{1-\mathcal{D}^{(0)}\pi^{(1)}} \tag{3.290}$$

The Green's function is also a function of the coupling constant strength η. Equation (3.288) is rewritten as

$$\Omega - \Omega_0 = -\frac{\nu}{2\beta} \sum_{iq_n} \int \frac{d^3q}{(2\pi)^3} \int_0^1 \frac{d\eta}{\eta} \pi^{(1)}(\eta, \mathbf{q}, iq_n) \mathcal{D}^{(1)}(\eta, \mathbf{q}, iq_n) \tag{3.291}$$

This expression is an approximation to Ω since it employs an approximate self-energy $\pi^{(1)}$. There is a theorem which we are going to state but not prove. The exact correction to the thermodynamic potential from electron–phonon interactions is

$$\Omega - \Omega_0 = -\frac{\nu}{2\beta} \sum_{iq_n} \int \frac{d^3q}{(2\pi)^3} \int_0^1 \frac{d\eta}{\eta} \pi(\eta, \mathbf{q}, iq_n) \mathcal{D}(\eta, \mathbf{q}, iq_n) \tag{3.292}$$

where $\pi(\eta, \mathbf{q}, iq_n)$ is the exact phonon self-energy and $\mathcal{D}(\eta, \mathbf{q}, iq_n)$ is the exact phonon Green's function evaluated with the exact self-energy. Both expressions are found as a function of η, and then one has to perform the integration over η. Proofs of this theorem are given in Abrikosov *et al.* (1963).

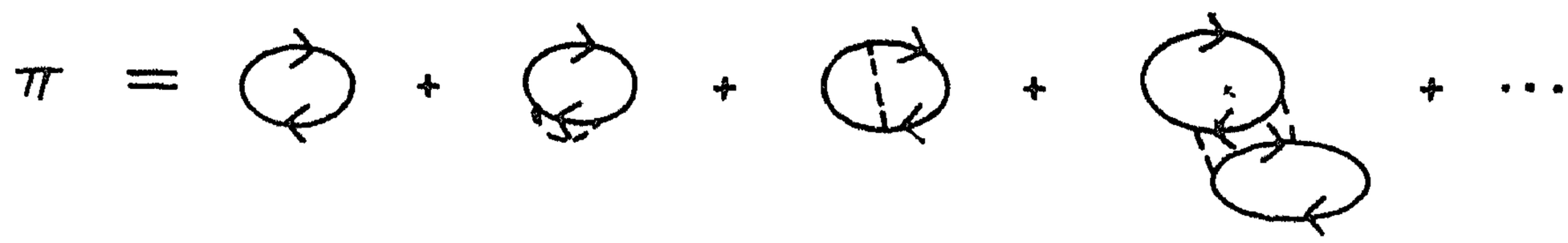


FIGURE 3.10

The exact phonon self-energy $\pi(\eta, \mathbf{q}, iq_n)$ is the sum of an infinite number of self-energy diagrams, some of which are shown in Fig. 3.10. There are bubbles with internal phonon lines and also several bubbles connected by more than one phonon line.

Of course, if electron–electron interactions are included, then the phonon self-energy can also contain internal Coulomb interactions, etc. The effect of Coulomb interaction on the thermodynamic potential should also be included. The general theorem (3.265) and (3.266) is true for all interactions: Coulomb, phonon, or others.

It is possible to combine the effects of the phonons and Coulomb interaction in a simple way. Consider electrons interacting by either or both interactions, and then the single-bubble diagram appears twice. As shown in Fig. 3.11, the two electron vertices can be connected by a single Coulomb line or a single phonon line. The sum of these contributions is

$$U_2 = \frac{\eta}{2} \sum_{\mathbf{q}, iq_n} P^{(1)}(\mathbf{q}, iq_n)[v_q + M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n)] \tag{3.293}$$

where $v_q = 4\pi e^2/q^2$. Similarly, there are four diagrams with two bubbles. The two bubbles interact by two Coulomb lines, two phonon lines, or a mix with one of each. This contribution is

$$U_4 = \frac{\eta^2}{4} \sum_{\mathbf{q}, iq_n} \{P^{(1)}(\mathbf{q}, iq_n)[v_q + M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n)]\}^2 \tag{3.294}$$

It seems desirable to introduce a combined interaction propagator, Coulomb plus phonon, which is

$$W^{(0)}(\mathbf{q}, iq_n) = v_q + M_{\mathbf{q}}^2 \mathcal{D}^{(0)}(\mathbf{q}, iq_n) \tag{3.295}$$

It obeys a Dyson equation, which is

$$W(\mathbf{q}, iq_n) = \frac{W^{(0)}(\mathbf{q}, iq_n)}{1 - W^{(0)}(\mathbf{q}, iq_n)P^{(1)}} \tag{3.296}$$

where $W(\mathbf{q}, iq_n)$ is the total self-energy diagram, which is the sum of the diagrams shown in Fig. 3.9. Of course, now the interaction lines in Fig. 3.9 refer to $W^{(0)}$, since they may be either

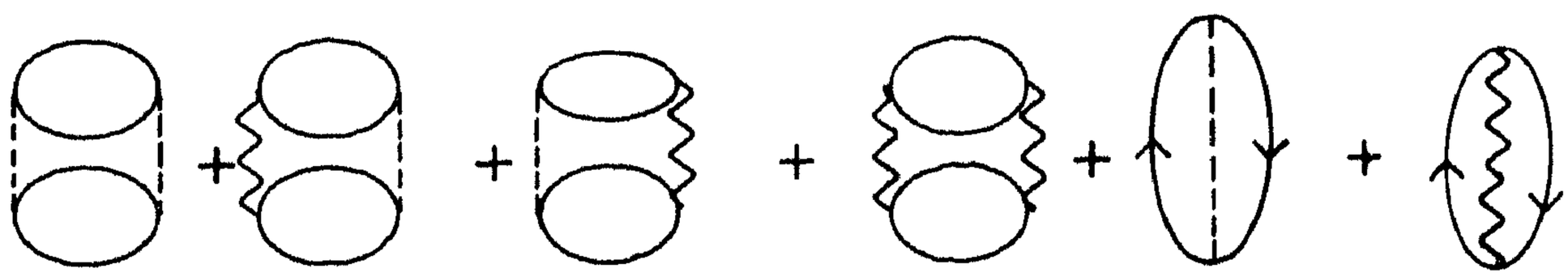


FIGURE 3.11

Coulomb or phonon. The generalization of (3.292) to include both Coulomb and phonon effects is

$$\Omega - \Omega_0 = -\frac{\nu}{2\beta}\sum_{iq_n}\int\frac{d^3q}{(2\pi)^3}\int_0^1\frac{d\eta}{\eta}\pi(\eta, \mathbf{q}, iq_n)W(\eta, \mathbf{q}, iq_n)$$
$$-\frac{\eta}{2}\int\frac{d^3q}{(2\pi)^3}v_q \tag{3.297}$$

The first term contains the Coulomb self-energy of an electron interacting with itself, and this unwanted contribution is subtracted out in the second term. The two interactions in (3.295) do not complete the possibilities. If the Hamiltonian has additional interactions due to impurities, ions, etc., then one has additional contributions which may contribute to the thermodynamic potential.

The expression (3.292) seems to favor the phonon self-energies. It is reasonable to ask what happened to the electron self-energies. The answer is that they are in there too. The diagrams under consideration can also be viewed as corrections to the electron self-energy. This viewpoint does not introduce an additional contribution but is another equivalent way of expressing the above answer. For example, the basic bubble diagram, shown in Fig. 3.7(a) and expressed in (3.272), may also be written as

$$U_2 = \frac{\lambda}{2}\sum_{ip_n,\mathbf{p}\sigma}\mathcal{G}^{(0)}(\mathbf{p}, ip_n)\Sigma^{(1)}(\mathbf{p}, ip_n) \tag{3.298}$$

where $\Sigma^{(1)}(\mathbf{p}, ip_n)$ is the electron self-energy from one-phonon processes, which was given earlier in (3.212):

$$\Sigma^{(1)}(\mathbf{p}, ip_n) = \frac{1}{\beta\nu}\sum_{\mathbf{q},iq_m}M_{\mathbf{q}}^2\mathcal{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip_n + iq_m)\mathcal{D}^{(0)}(\mathbf{q}, iq_m)$$

We repeat, for emphasis, that the above U_2 is not a different result than (3.280). It is the same result, expressed with different symbols; the same Green's functions have been grouped differently. Similarly, the exact answer is

$$\Omega - \Omega_0 = \frac{1}{\beta}\sum_{\mathbf{p},ip_n}\int_0^1\frac{d\eta}{\eta}\mathcal{G}(\eta, \mathbf{p}, ip_n)\Sigma(\eta, \mathbf{p}, ip_n) \tag{3.299}$$

where $\Sigma(\eta, \mathbf{p}, ip_n)$ is the exact self-energy of the electron and $\mathcal{G}(\eta, \mathbf{p}, ip_n)$ is the exact Green's function obtained from Dyson's equation with the exact self-energy. If the propagator W in (3.295) is used for all internal lines so that the electrons can interact via phonons or Coulomb interactions, then (3.299) is exactly equal to (3.297). The two results represent the same set of corrections. They are not different contributions but are the same contribution expressed in different notation.

These formulas are often used to calculate the ground state energy in the limit of $T \to 0$. There are two limits which are taken: $\nu \to \infty$ and $T \to 0$. In studying the electron gas, Kohn and Luttinger (1960) observed that the right answer was obtained if they were taken first in the limit of $\nu \to \infty$ and then $T \to 0$. The reverse order omits important terms, which are of the form

$$\beta n_F(\xi_{\mathbf{p}})[1 - n_F(\xi_{\mathbf{p}})] \tag{3.300}$$

If $\nu \neq \infty$, then all levels are discrete in the finite volume, and as $T \to 0$, these terms give zero since either n_F or $[1 - n_F]$ is zero. However, if one first takes $\nu \to \infty$, so that the levels are continuous, then the limit $T \to 0$ gives

$$\lim_{T\to 0} \beta n_F(\xi_{\mathbf{p}})[1 - n_F(\xi_{\mathbf{p}})] = \lim_{T\to 0} \frac{d}{d\xi_{\mathbf{p}}}[-n_F(\xi_{\mathbf{p}})] = \delta(\xi_{\mathbf{p}}) \tag{3.301}$$

There are many terms of this kind in the perturbation expansion for the ground state energy of the electron gas. They are called *dangerous diagrams*. They do not threaten your health, but they do threaten the accuracy of your answer.

It is illustrative to show a simple example of evaluating the thermodynamic potential. Assume that the only effect of the phonon self-energy is to change the unperturbed frequencies $\omega_{\mathbf{q}}^2$ to a new set of renormalized frequencies $\Omega_{\mathbf{q}}^2$. Since the Green's function is

$$D = \frac{2\omega_{\mathbf{q}}}{(i\omega_n)^2 - \omega_{\mathbf{q}}^2 - 2\omega_{\mathbf{q}}\pi(\mathbf{p}, i\omega_n)} = \frac{2\omega_{\mathbf{q}}}{(i\omega_n)^2 - \Omega_{\mathbf{q}}^2} \tag{3.302}$$

This form for D can be accomplished choosing

$$2\omega_{\mathbf{q}}\pi(\eta, \mathbf{p}, i\omega_n) = \eta(\Omega_{\mathbf{q}}^2 - \omega_{\mathbf{q}}^2) \tag{3.303}$$

The choice (3.303) is not the only possible one to renormalize the frequencies. It is the choice one gets by assuming that the change from $\omega_{\mathbf{q}}$ to $\Omega_{\mathbf{q}}$ is accomplished by the one-bubble polarization diagram given in (3.289). Then the coupling constant η enters as just a multiplicative factor, as shown in (3.303). If further self-energy diagrams are needed to get a good phonon self-energy π, then η would enter in a more complicated fashion. It is shown in Chapter 6 that the single-bubble approximation is often adequate, so that the present derivation applies to many systems.

Consider the steps necessary to evaluate (3.292), or perhaps just (3.288). The expression to be evaluated is

$$\Omega - \Omega_0 = -\frac{1}{2\beta}\sum_{\mathbf{q}, iq_n}\int_0^1 d\eta \frac{\Omega_{\mathbf{q}}^2 - \omega_{\mathbf{q}}^2}{(iq_n)^2 - \omega_{\mathbf{q}}^2 - \eta(\Omega_{\mathbf{q}}^2 - \omega_{\mathbf{q}}^2)} \tag{3.304}$$

First, introduce a frequency which depends on a coupling constant:

$$\Omega_\eta^2 = \omega_{\mathbf{q}}^2 + \eta(\Omega_{\mathbf{q}}^2 - \omega_{\mathbf{q}}^2) \tag{3.305}$$

The summation over Matsubara frequencies may be done by the techniques suggested in Sec. 3.5. Alternately, one might recognize the answer as

$$\frac{1}{\beta}\sum_{iq_n}\frac{1}{(iq_n)^2 - \Omega_\eta^2} = \frac{1}{2\Omega_\eta}\bar{\mathcal{D}}_\eta(\tau = 0) = -\frac{1}{2\Omega_\eta}[2n_B(\Omega_\eta) + 1] \tag{3.306}$$

where $\bar{\mathcal{D}}_\eta(\tau)$ is the Green's function which is the Fourier transform of

$$\bar{\mathcal{D}}_\eta(iq_n) = \frac{2\Omega_\eta}{(iq_n)^2 - \Omega_\eta^2} \tag{3.307}$$

The evaluation of (3.304) is now just the coupling constant integral,

$$\Omega - \Omega_0 = \frac{1}{2}\sum_{\mathbf{q}}(\Omega_{\mathbf{q}}^2 - \omega_{\mathbf{q}}^2)\int_0^1 \frac{d\eta}{2\Omega_\eta}\left(1 + \frac{2}{e^{\beta\Omega_\eta} - 1}\right) \tag{3.308}$$

which can be done exactly since $(\Omega^2 - \omega^2)d\eta/(2\Omega_\eta) = d\Omega_\eta$:

$$\Omega - \Omega_0 = \frac{1}{2}\sum_{\mathbf{q}}\left[\Omega_\eta + \frac{2}{\beta}\ln(1 - e^{-\beta\Omega_\eta})\right]_{\eta=0}^{\eta=1} \tag{3.309}$$

$$= \sum_{\mathbf{q}}\left[\frac{1}{2}(\Omega_{\mathbf{q}} - \omega_{\mathbf{q}}) + \frac{1}{\beta}\ln\left(\frac{1 - e^{-\beta\Omega_{\mathbf{q}}}}{1 - e^{-\beta\omega_{\mathbf{q}}}}\right)\right] \tag{3.310}$$

The right-hand side is the thermodynamic potential from the phonons at the new frequency $\Omega_{\mathbf{q}}$ minus the contribution from the phonons at the old frequencies $\omega_{\mathbf{q}}$. When this result is combined with (3.260) for Ω_0, the terms with $\omega_{\mathbf{q}}$ all cancel. The final answer is

$$\Omega = -\frac{2\nu}{\beta}\int\frac{d^3p}{(2\pi)^3}\ln(1 + e^{-\beta\xi_{\mathbf{p}}}) + \frac{\nu}{\beta}\int\frac{d^3q}{(2\pi)^3}[\tfrac{1}{2}\beta\Omega_{\mathbf{q}} + \ln(1 - e^{-\beta\Omega_{\mathbf{q}}})] \tag{3.311}$$

The thermodynamic potential is just the summation of the unperturbed electron contribution plus the contribution from the phonons at the new frequencies $\Omega_{\mathbf{q}}$. Of course, this form is expected from simple considerations. If the only effect of the electron–phonon interactions is to change the phonon frequencies to the new values $\Omega_{\mathbf{q}}$, then one must be able to solve the Hamiltonian exactly and write it as

$$H = \sum_{\mathbf{p}\sigma}\xi_{\mathbf{p}}C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} + \sum_{\mathbf{q}}\Omega_{\mathbf{q}}a^\dagger_{\mathbf{q}}a_{\mathbf{q}} \tag{3.312}$$

The thermodynamic potential for this simple Hamiltonian is just (3.311). Unfortunately, most corrections to the thermodynamic potential are not as easy to evaluate as this simple example.

3.6.2. Green's Functions

So far the linked cluster, or cumulant, expansion has only been used to evaluate the thermodynamic potential. One could use similar expansion techniques to evaluate other types of correlation functions, such as Green's functions (Brout and Carruthers, 1963). This method has been used with success on two problems. One is the polaron problem for a single particle in a band, studied by Dunn (1975). The other has been in treating X-ray processes, where the hole dynamics have been studied by a linked cluster method (Mahan, 1975). Here the hole has also been treated as a single-hole problem. It has not been applied to a many-polaron, or many-hole, problem. The procedure begins with the S-matrix expansion for a particle Green's function, given in (3.196):

$$\mathcal{G}(\mathbf{p}, \tau) = \sum_{n=0}^{\infty}\lambda^n W_n(\mathbf{p}, \tau)$$

$$W_n(\mathbf{p}, \tau) = -\frac{(-1)^n}{n!}\int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_{n0}\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau_1)\hat{V}(\tau_1)\cdots\hat{V}(\tau_n)\hat{C}_{\mathbf{p}\sigma}(0)\rangle \tag{3.313}$$

where the coupling parameter λ has been added to the expansion. In the cumulant expansion, it is assumed that the series can be regrouped as an exponential power series in λ:

$$\mathcal{G}(\mathbf{p}, \tau) = \mathcal{G}^{(0)}(\mathbf{p}, \tau)\exp\left[\sum_{n=1}^{\infty}\lambda^n F_n(\mathbf{p}, \tau)\right] \tag{3.314}$$

The functions $F_n(\mathbf{p}, \tau)$ are found by equating terms with like powers of λ in the series expansions. That is, write

$$\exp\left(\sum_{j=1} \lambda^j F_j\right) = 1 + \lambda F_1 + \lambda^2\left(F_2 + \frac{1}{2!}F_1^2\right) + \lambda^3\left(F_3 + F_1F_2 + \frac{1}{3!}F_1^3\right)\ldots \tag{3.315}$$

so that, term by term, we have the equalities

$$W_1 = \mathcal{G}^{(0)}F_1 \tag{3.316}$$

$$W_2 = \mathcal{G}^{(0)}\left(F_2 + \frac{1}{2!}F_1^2\right) \tag{3.317}$$

$$W_3 = \mathcal{G}^{(0)}\left(F_3 + F_1F_2 + \frac{1}{3!}F_1^3\right) \tag{3.318}$$

The functions $W_n(\mathbf{p}, \tau)$ are scalar functions which are calculated using Wick's theorems. The unknown functions $F_n(\mathbf{p}, \tau)$ can be evaluated in terms of W_n by solving the above equations (Brout and Carruthers, 1963):

$$F_1 = \frac{W_1}{\mathcal{G}^{(0)}} \tag{3.319}$$

$$F_2 = \frac{W_2}{\mathcal{G}^{(0)}} - \frac{1}{2!}F_1^2 \tag{3.320}$$

$$F_3 = \frac{W_3}{\mathcal{G}^{(0)}} - F_1F_2 - \frac{1}{3!}F_1^3 \tag{3.321}$$

The basic idea is very simple. It is just a different way of regrouping the series which results from the S matrix. It provides an alternative to using Dyson equations. Presumably, one gets the exact answer if all terms are taken and evaluated. However, since one can only get the exact answer in trivial problems, the method is really useful in problems where one can calculate a few terms in the perturbation expansion.

There are two possible difficulties with this procedure. The first is that the regrouping may not make mathematical sense. Nonsensical results are found in cases where $\mathcal{G}(\mathbf{p}, \tau)$ describes a Green's function of a particle in a system with N_e like particles, where $N_e \propto \nu$. In that case the terms F_n do not converge when put into the exponent. The offending terms are those which contribute to the total energy of the system and hence change the chemical potential. These energy terms are of order $O(\nu)$. The chemical potential is not changed by V in one-particle or one-hole problems, since the one particle does not contribute any energy term proportional to ν. The series in $F_n(\mathbf{p}, \tau)$ shows excellent convergence in these cases. These results will be discussed in Chapters 7 and 9. The other problem with this method is that, when it works, one has some fairly complicated functions of τ in the exponent. One is invariably forced to use a computer to perform the Fourier transforms in frequency space. This step is not really a difficulty—just some work!

The physical model implied by this regrouping will be discussed in Chapter 7. It is a systematic development of the Tomonaga (1947) model of pion emission in particle theory. Tomonaga assumed that all pion emissions from a particle were statistically independent—there was no correlation between successive emissions. His model corresponds to the

approximation of just using F_1 alone. However, if one uses $F_1 + F_2$, then F_2 puts in the correlation between pairs of emissions. Similarly, F_3 puts in correlations between three particle events, etc. The advantage of the method is that the Green's function includes many-particle emission processes, even when one keeps only a few terms in the series for F_n. Solid state problems do not have pions but other boson excitations such as phonons or the boson-like excitations of the electron gas.

3.7. REAL-TIME GREEN'S FUNCTIONS

Six different Green's function of time were introduced in Chapter 2. Two of them, the retarded and advanced functions, have been discussed at nonzero temperature. The other four are discussed here. The usefulness of the real-time functions is in the treatment of nonequilibrium phenomena. Nonequilibrium transport theory using real-time Green's function is treated in Chapter 8. The present section discusses their properties at nonzero temperature.

The Matsubara method is unsuitable for nonequilibrium since there is no thermodynamic basis for temperature for a system out of equilibrium. The entire Matsubara method is based upon temperature, and no method has been found so far for extending it to nonequilibrium processes.

The real-time functions at nonzero temperature have formal definitions very similar to those at zero temperature. Comparing with (2.144), the Green's functions for electrons of momentum $\mathbf{p}$ are

$$\begin{aligned}
G^{<}(\mathbf{p}, t_1, t_2) &= i\langle C^{\dagger}_{\mathbf{p}\sigma}(t_2) C_{\mathbf{p}\sigma}(t_1)\rangle \\
G^{>}(\mathbf{p}, t_1, t_2) &= -i\langle C_{\mathbf{p}\sigma}(t_1) C^{\dagger}_{\mathbf{p}\sigma}(t_2)\rangle \\
G_t(\mathbf{p}, t_1, t_2) &= \Theta(t_1 - t_2) G^{>}(\mathbf{p}, t_1, t_2) + \Theta(t_2 - t_1) G^{<}(\mathbf{p}, t_1, t_2) \\
G_{\bar{t}}(\mathbf{p}, t_1, t_2) &= \Theta(t_2 - t_1) G^{>}(\mathbf{p}, t_1, t_2) + \Theta(t_1 - t_2) G^{<}(\mathbf{p}, t_1, t_2)
\end{aligned} \tag{3.322}$$

These definitions appear to be identical to those at zero temperature. The important difference at nonzero temperatures is that the brackets $\langle \cdots \rangle$ have a different meaning. They have several different meanings, depending upon the circumstance:

Case 1: At zero temperature, and in equilibrium, the brackets $\langle \cdots \rangle$ denote the ground state of the interacting system.

Case 2: At nonzero temperature, and in equilibrium, the brackets denote the thermodynamic average as given in (3.25) or (3.82).

Case 3: When not in equilibrium, the brackets denote an average over the accessible phase space. However, the available phase space depends upon the recent history of the system, and kinetic constraints such as energy conservation. The meaning of the brackets is poorly understood for systems out of equilibrium.

Another feature of (3.322) is that the Green's functions are not expressed on the left as functions of $(t_1 - t_2)$ but of (t_1, t_2) separately. In equilibrium, they are functions of $(t_1 - t_2)$, but that may not be true for systems out of equilibrium.

In thermal equilibrium, the real-time functions each have a simple relation to the retarded function. Each is found easily from G_{ret}. Since G_{ret} is found easily from the Matsubara

function, the easiest way to find the real-time functions is by first finding the Matsubara function of complex frequency. These statements apply only in thermal equilibrium.

In thermal equilibrium the following derivation establishes the relationships among the real-time Green's functions. Using the states $|n\rangle$ and $|m\rangle$, which are exact eigenstates of the Hamiltonian, two of the real-time functions in (3.322) can be written as

$$G^{<}(\mathbf{p}, t_1 - t_2) = i \sum_{nm} e^{-\beta E_m} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 e^{it(E_n - E_m)(t_1 - t_2)} \tag{3.323}$$

$$G^{>}(\mathbf{p}, t_1 - t_2) = -i \sum_{nm} e^{-\beta E_n} |\langle n|C_{\mathbf{p}\sigma}|m\rangle|^2 e^{it(E_n - E_m)(t_1 - t_2)} \tag{3.324}$$

The time dependence is Fourier transformed to ω, which produces the delta function $2\pi\delta(\omega + E_n - E_m)$ in each term. These terms have many of the same factors that occur in the spectral function in (3.120). They differ mainly in the thermal factors $\exp(-\beta E_{n,m})$. They can be made to be identical by utilizing some algebraic relationships such as

$$\begin{aligned} e^{-\beta E_m}\delta(\omega + E_n - E_m) &= e^{-\beta E_m}[1 + e^{\beta\omega}]n_F(\omega)\delta(\omega + E_n - E_m) \\ &= [e^{-\beta E_m} + e^{-\beta E_n}]n_F(\omega)\delta(\omega + E_n - E_m) \end{aligned}$$

In the first line the factor of $n_F(\omega)$ has been added to both numerator and denominator. In the second line, one of these factors has been changed to the sum of two exponential factors by using $E_m - \omega = E_n$. The part in brackets is exactly the combination that appears in the spectral function:

$$G^{<}(\mathbf{p}, \omega) = i n_F(\omega) A(\mathbf{p}, \omega) \tag{3.325}$$

The other real-time functions can be derived using the same technique:

$$G^{>}(\mathbf{p}, \omega) = -i[1 - n_F(\omega)]A(\mathbf{p}, \omega) \tag{3.326}$$

$$G_t(\mathbf{p}, \omega) = [1 - n_F(\omega)]G_{\text{ret}}(\mathbf{p}, \omega) + n_F(\omega)G_{\text{adv}}(\mathbf{p}, \omega) \tag{3.327}$$

$$G_{\bar{t}}(\mathbf{p}, \omega) = -[1 - n_F(\omega)]G_{\text{adv}}(\mathbf{p}, \omega) - n_F(\omega)G_{\text{ret}}(\mathbf{p}, \omega) \tag{3.328}$$

These expressions are identical to (2.160). In equilibrium the real-time Green's functions are found easily from the retarded and advanced functions. The latter two are found easily from the Matsubara functions. In equilibrium there is no need to set up a separate formalism for the real-time functions.

The primary usefulness of the real-time Green's functions is in the theory of nonequilibrium phenomena. Some of these ideas and techniques are introduced here. The first step is to derive Dyson's equation. In fact the result (2.157) at zero temperature is also correct for nonzero temperature. The matrices have the definitions that are given in (2.156). There is no reason to rederive the same equation. Instead it seems appropriate to comment on the differences in the derivation between zero and nonzero temperatures.

Equation (2.157) is generally regarded as being the correct form for Dyson's equation, even for systems out of equilibrium. However, we have never seen a satisfactory proof for nonequilibrium systems. For systems in equilibrium, one can derive (2.157) at nonzero temperature by starting from Dyson's equation for the Matsubara Green's functions. Treating

τ as a complex variable, one can deform the contour of integration, and end up with (2.157) (Langreth, 1976).

For nonequilibrium, the Matsubara functions are obviously an invalid starting point. Instead, one can try to expand the S matrix in real time. Then one encounters the problems mentioned in Sec. 3.1 regarding expanding two S matrices: one for $\exp(-\beta H)$ and another one for $\exp(\pm itH)$. All derivations ignore the former. Including only the S matrix of time, one again finds (2.157). It appears correct to omit the S-matrix expansion of $\exp(-\beta H)$. This point is discussed in the next section. Equation (2.157) is used for Dyson's equation for nonzero temperature, for equilibrium and nonequilibrium, since there is nothing else available.

Nonequilibrium theory usually proceeds by first deriving an equation of motion for the Green's function, which is similar to a Boltzmann equation. This equation is usually derived for systems in equilibrium, or slightly out of equilibrium. Then the equation is applied to systems far from equilibrium.

The first step is to find an equation of motion for the interacting Green's function when it is not in equilibrium. Such an equation can be found from (2.157) by operating by $(i\partial/\partial t - H)$ on both sides of the equation. First one needs to know the time evolution of the noninteracting Green's functions. This behavior is deduced easily from (2.144) and (2.148). The time derivatives can be expressed compactly in the matrix notation. Remember that $d\Theta(t)/dt = \delta(t)$.

$$\left(i\frac{\partial}{\partial t} - \varepsilon_{\mathbf{k}}\right)\tilde{G}_0(\mathbf{k}, t) = \delta(t)\tilde{I} \tag{3.329}$$

$$\left(i\frac{\partial}{\partial t} - H_0(x)\right)\tilde{G}_0(x) = \delta^4(x)\tilde{I} \tag{3.330}$$

On the right-hand side of (2.157), this operator acts only upon $\tilde{G}_0$ for which the above result is used to find

$$\left(i\frac{\partial}{\partial t} - H_0(x)\right)\tilde{G}(x_1, x_2) = \delta^4(x_1 - x_2)\tilde{I} + \int dx_3 \tilde{\Sigma}(x_1, x_3)\tilde{G}(x_3, x_2) \tag{3.331}$$

This formula provides the equation of motion for the interacting Green's function. It is the basis for the nonequilibrium theory of interacting systems.

The structure of this equation is interesting. On the left-hand side, only the noninteracting terms are contained in the Hamiltonian H_0. The contribution from the interactions is provided by the self-energy functions on the right-hand side.

So far, the equation of motion has been derived for the variable x_1 in $\tilde{G}(x_1, x_2)$. It is useful to have an equation of motion for the other variable x_2. Since the definition (2.144) of the Green's functions contains the conjugate wave function $\psi^\dagger(x_2)$, the equation of motion on this variable is the complex conjugate of Schrödinger's equation. Furthermore, Dyson's equation can be written in an alternate form. Instead of (2.157), an alternate form is to have $\tilde{G}_0$ on the right in the interaction terms:

$$\tilde{G}(x_1, x_2) = \tilde{G}_0(x_1 - x_2) + \int dx_3 \int dx_4 \tilde{G}(x_1, x_3)\tilde{\Sigma}(x_3, x_4)\tilde{G}_0(x_4 - x_2)$$

The equation of motion on the x_2 variable, when using this form for Dyson's equation, still acts only upon $\tilde{G}_0$ on the right and produces delta functions. These steps produce the alternate equation of motion for the Green's function:

$$\left(-i\frac{\partial}{\partial t} - H_0(\mathbf{r}_2, -\mathbf{p}_2)\right)\tilde{G}(x_1, x_2) = \delta^4(x_1 - x_2)\tilde{I} + \int dx_3 \tilde{G}(x_1, x_3)\tilde{\Sigma}(x_3, x_2) \tag{3.332}$$

The sign change on $\mathbf{p}_2$ comes from the fact that the left-hand side of the above equation is the complex conjugate of Schrödinger's equation. The operator $\mathbf{p} = -i\nabla$ changes sign under complex conjugation. This behavior is different from the Hermitian conjugate, where H_0 acts to the left on $\psi^\dagger(x_2)$ and $\mathbf{p}$ does not change sign. These two equations of motion will be used in Chapter 8 to develop a quantum Boltzmann equation for nonequilibrium phenomena. The quantum Boltzmann equation is based upon a many-body distribution function first suggested by Wigner (1932).

3.7.1. Wigner Distribution Function

The traditional Boltzmann equation is expressed in terms of the distribution function f ($\mathbf{r}$, $\mathbf{v}$, t). The three variables are position, $\mathbf{r}$, velocity, $\mathbf{v}$, and time, t. The point of view is semiclassical since it is assumed that the position and velocity (momentum) of the particle can be defined simultaneously. In order to use this distribution for quantum systems, it is necessary to perform some type of averaging in order to remove effects due to the uncertainty principle.

If quantum effects are important, it is necessary to introduce another variable into the distribution function. This variable could either be energy E or equivalently a frequency $\omega = E$. The resulting distribution function is often called a Wigner distribution function f ($\mathbf{k}$, ω; $\mathbf{r}$, t). The velocity $\mathbf{v}$ has been changed to the wave vector $\mathbf{k} = m\mathbf{v}$. The distribution function is derived from the Green's function $G^<(x_1, x_2)$ defined in Eq. (2.144). The Wigner distribution function (WDF) is derived using the following series of steps. First go to a center-of-mass coordinate system:

$$(\mathbf{R}, T) = \tfrac{1}{2}(x_1 + x_2) \tag{3.333}$$

$$(\mathbf{r}, t) = x_1 - x_2 \tag{3.334}$$

Note that T means center-of-mass time, rather than temperature or time-ordering. The symbol β is used for the inverse of Boltzmann's constant times temperature. The notation on the Green's function is altered to these center-of-mass coordinates:

$$G^<(\mathbf{r}, t; \mathbf{R}, T) = i\langle\psi^\dagger(\mathbf{R} - \tfrac{1}{2}\mathbf{r}, T - \tfrac{1}{2}t)\psi(\mathbf{R} + \tfrac{1}{2}\mathbf{r}, T + \tfrac{1}{2}t)\rangle \tag{3.335}$$

The next step is to Fourier transform the relative variables ($\mathbf{r}$, t) into ($\mathbf{k}$, ω):

$$G^<(\mathbf{k}, \omega; \mathbf{R}, T) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \int dt e^{i\omega t} G^<(\mathbf{r}, t; \mathbf{R}, T) \tag{3.336}$$

The relation to the WDF is quite simple:

$$f(\mathbf{k}, \omega; \mathbf{R}, T) = -iG^{<}(\mathbf{k}, \omega; \mathbf{R}, T) \tag{3.337}$$

Regard this assertion as the definition of $f(\mathbf{k}, \omega; \mathbf{R}, T)$. This choice is made reasonable by showing that various moments of f provide the macroscopic quantities of particle density $n(\mathbf{R}, T)$, particle current $j_\mu(\mathbf{R}, T)$, and energy density $n_E(\mathbf{R}, T)$:

$$\begin{aligned} n(\mathbf{R}, T) &= \int \frac{d^3k}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} f(\mathbf{k}, \omega; \mathbf{R}, T) = \langle \psi^\dagger(\mathbf{R}, T)\psi(\mathbf{R}, T)\rangle \\ j(\mathbf{R}, T) &= \int \frac{d^3k}{(2\pi)^3} \frac{\mathbf{k}}{m} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} f(\mathbf{k}, \omega; \mathbf{R}, T) \end{aligned} \tag{3.338}$$

$$\begin{aligned} n_E(\mathbf{R}, T) &= \int \frac{d^3k}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \omega f(\mathbf{k}, \omega; \mathbf{R}, T) \\ &= i\left[\frac{\partial}{\partial t} \langle \psi^\dagger(\mathbf{R}, T - \tfrac{1}{2}t)\psi(\mathbf{R}, T + \tfrac{1}{2}t)\rangle\right]_{t=0} \\ &= \langle \psi^\dagger(\mathbf{R}, T) H \psi(\mathbf{R}, T)\rangle \end{aligned} \tag{3.339}$$

The technique for solving nonequilibrium problems is very simple. The equation of motion in (3.331) or (3.332) for $G^{<}(\mathbf{k}, \omega; \mathbf{R}, T)$ is just the quantum Boltzmann equation (QBE). This equation is then solved, which yields directly the Wigner distribution function $f(\mathbf{k}, \omega; \mathbf{R}, T)$. Macroscopic variables such as current and density are found by taking the above integrals over d^3k and $d\omega$. This technique is also useful in systems that are not homogeneous because there is a slowly varying potential $V(\mathbf{R})$.

The above expressions are useful for nonequilibrium situations. For example, the answer is zero if the equilibrium function (3.325) for $G^{<}$ is used to calculate the current j_μ in (3.338). That is the right answer. A system carrying current can be in steady state but not in equilibrium. In order to obtain a nonzero current it is necessary to solve for the nonequilibrium Green's function $G^{<}$ when the Hamiltonian contains an electric field. This procedure is described in Chapter 8.

The semiclassical Boltzmann distribution function $f(\mathbf{R}, \mathbf{v}, T)$ is found by taking the frequency integral of the WDF:

$$f(\mathbf{R}, \mathbf{v}, T) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} f(m\mathbf{v}, \omega; \mathbf{R}, T) \tag{3.340}$$

This method of deriving the semiclassical Boltzmann equation is an alternative to the usual technique of coarse grain averaging.

The Green's function $G^{<}$ plays a central role in the nonequilibrium theory. In order to solve it one also needs to know the retarded function.

These two Green's functions are the most important. The matrix equation in (3.331) and (3.332) is untangled to present the individual equation of motion for these separate Green's functions.

$$\left[i\frac{\partial}{\partial t_1} - H_0(x_1)\right]G^<(x_1, x_2) = \int dx_3[\Sigma_t(x_1, x_3)G^<(x_3, x_2) - \Sigma^<(x_1, x_3)G_{\bar{t}}(x_3, x_2)]$$

$$\left[i\frac{\partial}{\partial t_1} - H_0(x_1)\right]G_{\text{ret}}(x_1, x_2) = \delta^4(x_1 - x_2) + \int dx_3\Sigma_{\text{ret}}(x_1, x_3)G_{\text{ret}}(x_3, x_2) \tag{3.341}$$

$$\left[-i\frac{\partial}{\partial t_2} - H_0(x_2)^*\right]G^<(x_1, x_2) = \int dx_3[G_t(x_1, x_3)\Sigma^<(x_3, x_2) - G^<(x_1, x_3)\Sigma_{\bar{t}}(x_3, x_2)]$$

$$\left[-i\frac{\partial}{\partial t_2} - H_0(x_2)^*\right]G_{\text{ret}}(x_1, x_2) = \delta^4(x_1 - x_2) + \int dx_3\, G_{\text{ret}}(x_1, x_3)\Sigma_{\text{ret}}(x_3, x_2) \tag{3.342}$$

The equations are the starting point for the derivation of the quantum Boltzmann equation.

The WDF is an elegant and useful formalism. It does have some liabilities. If one constructs the WDF for any nontrivial Hamiltonian, then $f(\mathbf{r}, \omega; \mathbf{R}, T)$ has regions where it is negative. The function is not positive definite, which means that it cannot be interpreted as a probability density. This feature can be shown by a simple example. Consider in one dimension a particle in a box of length L. The eigenvalues and eigenfunctions are

$$\phi_n(x) = \sqrt{\frac{2}{L}}\sin(k_n x), \quad \varepsilon_n = \varepsilon n^2 \tag{3.343}$$

$$k_n = \frac{n\pi}{L}, \quad \varepsilon = \frac{\pi^2}{2mL^2} \tag{3.344}$$

Using the representation in (2.147), and setting $t = t_1 - t_2$, gives

$$G^<(x_1, x_2, t) = \frac{2i}{L}\sum_n n_F(\varepsilon_n)\sin(k_n x_1)\sin(k_n x_2)e^{-it\varepsilon n^2} \tag{3.345}$$

Here $x_{1,2}$ are not four-vectors but one-dimensional coordinates of position. The product of two sine functions can be written as the difference of two cosine functions. Taking the Fourier transform of time gives the function of frequency. Since the system is in equilibrium, write $G^< = in_F(\omega)A$. In the center-of-mass notation

$$f(x, \omega; X) = n_F(\omega)A(x, \omega; X) \tag{3.346}$$

$$A(x, \omega; X) = \frac{2\pi}{L}\sum_n[\cos(k_n x) - \cos(2k_n X)]\delta(\omega - \varepsilon n^2) \tag{3.347}$$

where $x = x_1 - x_2$ is relative position and $X = (x_1 + x_2)/2$ is center-of-mass position. The function $A(x, \omega; X)$ is a sum of delta functions. The amplitudes of the delta functions have variable sign, depending upon the values of x and X. Neither A nor f are positive at all points.

For many Hamiltonians of interest the function f is not positive definite. That creates philosophical problems if f is interpreted as a probability function. Do not interpret f this way. Instead, f is regarded as one step in the calculation. It is never the last step, since it is used to calculate other quantities that can be measured. Physically sensible results are always found for measurable quantities such as the particle density and current. No problems are encountered as long as one avoids interpreting f as a probability density.

3.8. KUBO FORMULA FOR ELECTRICAL CONDUCTIVITY

Many experiments in condensed matter physics measure the linear response to an external perturbation. The experimentalists may put the sample in a magnetic field, electric field, optical field, temperature gradient, or pressure field and measure the magnetization, electrical current, light absorption, or whatever. Linear response means that the signal is directly proportional to the intensity of the external perturbation. Usually the assumption of linear response is valid at low magnitude of perturbing field.

Kubo formulas are the name applied to the correlation function which describes the linear response. There are many of them, since there are many possible perturbations and many linear responses for each perturbation. Formulas of this type were first proposed by Green (1952, 1954) for transport in liquids. Kubo (1957, 1959) first derived the equations for electrical conductivity in solids and in the form which will be used here. His derivation and result will be given below. Other Kubo formulas will be derived in this section. In fact, several derivations are given of the formula for the electrical conductivity in order to familiarize the reader with different types of derivations in the literature. One finds formulas which appear to be quite different but are really identical. It is also demonstrated that the longitudinal and transverse conductivities are identical at long wavelength.

In electrical conduction, a time-dependent external electric field,

$$E_\alpha^{(\text{ext})}(\mathbf{r}, t) = \Xi_\alpha^{(\text{ext})} e^{i\mathbf{q}\cdot\mathbf{r} - i\omega t} \tag{3.348}$$

is applied to the solid, where $\alpha = x, y, z$ are the directions in space. In linear response, the induced current is proportional to the applied electric field

$$J_\alpha(\mathbf{r}, t) = \sum_\beta \sigma'_{\alpha\beta}(\mathbf{q}, \omega) \Xi_\beta^{(\text{ext})} e^{i\mathbf{q}\cdot\mathbf{r} - i\omega t} \tag{3.349}$$

Although the symbol σ' has the appearance and dimensions of a conductivity, it is not the appropriate conductivity. Instead, the conductivity is needed which is the response to the total electric field in the solid. The applied or external field $E_\alpha^{(\text{ext})}$ induces currents which in turn make other electric fields. The summation of all these fields is the total electric field, which is called $E_\alpha(\mathbf{r}, t)$. The conductivity σ is the one which responds to the actual electric field in the solid:

$$J_\alpha(\mathbf{r}, t) = \sum_\beta \sigma_{\alpha\beta}(\mathbf{q}, \omega) E_\beta(\mathbf{r}, t) \tag{3.350}$$

$$E_\alpha(\mathbf{r}, t) = \Xi_\alpha \ \exp[i(\mathbf{q} \cdot \mathbf{r} - \omega t)] \tag{3.351}$$

Take Eq. (3.350) as the fundamental definition of the microscopic conductivity. It also introduces the coordinated problem of determining the proportionality between the external electric field $E^{(\text{ext})}$ and the total internal electric field $E_\alpha(\mathbf{r}, t)$. For a static, longitudinal electric field, the two electric fields are related by the continuity of the normal component of $\mathbf{D}$ at the surface. A similar relation for transverse fields is derived below. Equation (3.350) is correct for a homogeneous material. Actual solids are not homogeneous, although crystals are periodic. However, (3.350) implies the spacetime response may be written as

$$J_\alpha(\mathbf{r}, t) = \int d^3r' \int_{-\infty}^{t} dt' \sigma_{\alpha\beta}(\mathbf{r} - \mathbf{r}', t - t') E_\beta(\mathbf{r}', t') \tag{3.352}$$

where repeated indices imply summation. This equation is the Fourier transform of (3.350). It assumes that the current response of a material at $\mathbf{r}$ is only a function of the separation $(\mathbf{r} - \mathbf{r}')$ from the external electric field at $\mathbf{r}'$. This assumption is incorrect on an atomic scale. It is certainly not valid in the atomic-like core states of a solid or in any case where the separation $(\mathbf{r} - \mathbf{r}')$ is a few angstroms. A rigorous formulation would have the conductivity as a function of $\mathbf{r}$ and $\mathbf{r}'$ separately, $\sigma(\mathbf{r}, \mathbf{r}'; t - t')$. In solids, it is permissible to use (3.350) only when it is understood that the current is to be averaged over many unit cells of the solid. Usually it is applied when $\mathbf{q}$ is small and long-wavelength excitations are being studied.

Quite often our interest is in the dc conductivity, which is obtained by taking the limits $\mathbf{q} \to 0$ and $\omega \to 0$ in that order. Then the conductivity is only real. The Kubo formula is derived assuming that only a single frequency is perturbing the system and that $\sigma_{\alpha\beta}(\mathbf{q}, \omega)$ is the response to this single frequency. Actually, using a single frequency is not a restriction. It is assumed that the system is linear and perturbations at different frequencies act independently. The total current is then the summation of the responses at different frequencies.

The Hamiltonian for the system is taken to have the form

$$H + H' \tag{3.353}$$

The term H' contains the interaction between the total electric field and the particles of the system. Equation (2.161) is used as the basic form of the interaction between electromagnetic fields and charges. The electric field is expressed as a vector potential, so that

$$H' = -\frac{1}{c} \int d^3r j_\alpha(\mathbf{r}) A_\alpha(\mathbf{r}, t) \tag{3.354}$$

$$\frac{1}{c} A_\alpha(\mathbf{r}, t) = \frac{-i}{\omega} E_\alpha(\mathbf{r}, t) \tag{3.355}$$

Select the Coulomb gauge, where $\nabla \cdot \mathbf{A} = 0$, as explained in Sec. 2.10. Also, the electric and vector potentials are taken to be transverse, so the scalar potential ϕ is set equal to zero. The terms in (2.161) with A^2 are dropped, since their effects are nonlinear in the electric field. The term H contains all the other terms and interactions in the solid or liquid. There are interactions such as electron–electron, electron–phonon, spin–spin, with impurities, etc. The goal is to calculate the electrical conductivity when all these other interactions are present.

The current operator in (3.354) was discussed in Chapter 1. It has the form of

$$j_\alpha(\mathbf{r}) = \frac{1}{2m} \sum_i e_i [\mathbf{p}_{i\alpha} \delta(\mathbf{r} - \mathbf{r}_i) + \delta(\mathbf{r} - \mathbf{r}_i) \mathbf{p}_{i\alpha}] \tag{3.356}$$

The notation is more compact if the dr integral is evaluated in (3.354), which gives H' in terms of the Fourier transform of the current operator:

$$H' = \frac{i}{\omega} j_\alpha(\mathbf{q}) \Xi_\alpha e^{-i\omega t} \tag{3.357}$$

$$j_\alpha(\mathbf{q}) = \frac{1}{2m} \sum_i e_i [\mathbf{p}_{i\alpha} e^{i\mathbf{q}\cdot\mathbf{r}_i} + e^{i\mathbf{q}\cdot\mathbf{r}_i} \mathbf{p}_{i\alpha}] \tag{3.358}$$

In terms of creation and destruction operators, the current operator is conventionally written as

$$j_\alpha(\mathbf{q}) = \sum_{\lambda\delta} p_\alpha^{(\lambda\delta)} C_\lambda^\dagger C_\delta \tag{3.359}$$

where λ, δ are the states associated with some unperturbed Hamiltonian H_0 which is chosen as the basis for the perturbation expansion. Several of these possibilities were discussed in Chapter 1. A distinction is made between the current operator j_α in (3.358) and the induced current J_α in (3.350). The operator j_α is used in the Hamiltonian, while J_α is the actual current measured by the experimentalist. The measured value of the current is the average value for the velocity of the particles in the system, which is taken as the summation over all the particle velocities divided by the volume:

$$J_\alpha(\mathbf{r}, t) = \frac{e}{\mathrm{v}} \left\langle \sum_i v_{i\alpha} \delta(\mathbf{r} - \mathbf{r}_i) \right\rangle = \frac{e}{\mathrm{v}} \sum_i \langle v_{i\alpha} \rangle \tag{3.360}$$

When quantizing the particle velocity in (1.410), the velocity is the momentum minus the vector potential:

$$\mathbf{v}_i = \frac{1}{m} \left[\mathbf{p}_i - \frac{e}{c} \mathbf{A}(\mathbf{r}_i) \right] \tag{3.361}$$

$$J_\alpha = \frac{e}{m\mathrm{v}} \sum_i \langle p_{i\alpha} \rangle - \frac{e^2}{mc\mathrm{v}} \sum_i A_\alpha(\mathbf{r}_i) \tag{3.362}$$

The momentum operator $\mathbf{p}_i$ is proportional to the current operator, $\mathbf{j} = e\mathbf{p}_i/m$. The last term uses the relationship (3.354) between the vector potential and the electric field. For this term, the summation over particles divided by volume is replaced by the density, which is appropriate for a long-wavelength disturbance:

$$J_\alpha(\mathbf{r}, t) = \langle j_\alpha(\mathbf{r}, t) \rangle + i \frac{n_0 e^2}{m\omega} E_\alpha(\mathbf{r}, t) \tag{3.363}$$

The second term in the current is proportional to the electric field and the first term given by the expectation value of the local current operator. The latter term is also proportional to the electric field. Later it is shown that the constant of proportionality is given by the Kubo formula.

These two terms are called $\mathbf{J}^{(1)}$ and $\mathbf{J}^{(2)}$:

$$\mathbf{J} = \mathbf{J}^{(1)} + \mathbf{J}^{(2)} \tag{3.364}$$

$$\mathbf{J}^{(1)} = \frac{i n_0 e^2}{m\omega} \mathbf{E}(\mathbf{r}, t) \tag{3.365}$$

$$\mathbf{J}^{(2)} = \langle \mathbf{j}(\mathbf{r}, t) \rangle \tag{3.366}$$

The next step is to derive $\mathbf{J}^{(2)}$, which is the derivation of the Kubo formula.

3.8.1. Transverse Fields, Zero Temperature

The following derivation of the Kubo formula is valid at zero temperature. The first step is to consider the expectation value of the current operator as a function of time:

$$J_\alpha^{(2)}(\mathbf{r}, t) = \langle\psi'|e^{i(H+H')t} j_\alpha(\mathbf{r}) e^{-i(H+H')t}|\psi'\rangle \tag{3.367}$$

The Heisenberg representation is used, as explained in Sec. 2.2. Next go to the interaction representation, where H' is treated as the perturbation.

$$e^{-i(H+H')t} = e^{-itH} U(t) \tag{3.368}$$

$$U(t) = e^{itH} e^{-i(H+H')t} \tag{3.369}$$

$$J_\alpha^{(2)}(\mathbf{r}, t) = \langle\psi'|U^\dagger(t) e^{itH} j_\alpha(\mathbf{r}) e^{-itH} U(t)|\psi'\rangle \tag{3.370}$$

The operator $U(t)$ was defined earlier in (2.14), and a formal solution was derived in (2.27):

$$U(t) = T\exp\left[-i\int_0^t dt' H'(t')\right] \tag{3.371}$$

where the usual definitions are

$$H'(t) = e^{itH} H' e^{-itH} \tag{3.372}$$

$$j(t) = e^{itH} j e^{-itH}, \text{ etc.} \tag{3.373}$$

The wave function $|\psi'\rangle$ in (3.367) is the Schrödinger wave function at $t = 0$ for an interacting system with both $H + H'$ as the Hamiltonian. As discussed in Sec. 3.1, the wave function $|\psi\rangle$ is appropriate when H' is absent from the system. The relationship between the two wave functions is [see (2.36)]

$$|\psi'\rangle = T\exp\left[-i\int_{-\infty}^0 dt' H'(t')\right]|\psi\rangle \tag{3.374}$$

The time development of the system is given by combining these results:

$$U(t)|\psi'\rangle = T\exp\left[-i\int_{-\infty}^t dt' H'(t')\right]|\psi\rangle \tag{3.375}$$

$$= S(t, -\infty)|\psi\rangle \tag{3.376}$$

The expectation value of the current is now

$$J_\alpha^{(2)} = \langle\psi|S^\dagger(t, -\infty) j_\alpha(\mathbf{r}, t) S(t, -\infty)|\psi\rangle \tag{3.377}$$

An exact solution of this equation is not needed for the derivation of the Kubo formula. Only terms linear in the electric field E_α are desired, which requires us to only keep terms linear in H'. It is sufficient, for linear response, to keep the terms

$$S(t, -\infty)|\psi) = \left[1 - i\int_{-\infty}^t dt' H'(t')\right]|\psi\rangle + O(H')^2 \tag{3.378}$$

The Hermitian conjugate of this wave function is

$$\langle\psi|S^\dagger(t, -\infty) = \langle\psi|\left[1 + i\int_{-\infty}^t dt' H'(t')\right] + O(H')^2 \tag{3.379}$$

From (3.377), the expectation value of the current operator is

$$J_\alpha^{(2)}(\mathbf{r},t) = \langle\psi|\left[1 + i\int_{-\infty}^{t} dt' H'(t')\right] j_\alpha(\mathbf{r},t)\left[1 - i\int_{-\infty}^{t} dt' H'(t')\right]|\psi\rangle$$
$$= \langle\psi|\left[j_\alpha(\mathbf{r},t) - i\int_{-\infty}^{t} dt'[j_\alpha(\mathbf{r},t)H'(t') - H'(t')j_\alpha(\mathbf{r},t)]\right]|\psi\rangle$$

The first term is assumed to vanish,

$$\langle\psi|j_\alpha(\mathbf{r},t)|\psi\rangle = 0 \tag{3.380}$$

since there is usually no current in the solid in the absence of an electric field or something equivalent such as a time-varying magnetic field. The first nonzero term is the one linear in H', which is the important contribution. There are two terms, which can be expressed as a commutator:

$$J_\alpha^{(2)}(\mathbf{r},t) = -i\int_{-\infty}^{t} dt' \langle\psi|[j_\alpha(\mathbf{r},t), H'(t')]|\psi\rangle \tag{3.381}$$

The goal is to derive Eq. (3.350). A factor of $E_\beta(\mathbf{r}', t')$ must be removed from the integrand. With H' given in (3.358), the integrand has the factors

$$[j_\alpha(\mathbf{r},t), H'(t')] = \frac{i}{\omega}\Xi_\beta e^{-i\omega t'}[j_\alpha(\mathbf{r},t), j_\beta(\mathbf{q},t')]$$
$$= \frac{i}{\omega}E_\beta(\mathbf{r},t)e^{-i\mathbf{q}\cdot\mathbf{r}}e^{i\omega(t-t')}[j_\alpha(\mathbf{r},t), j_\beta(\mathbf{q},t')] \tag{3.382}$$

which are regrouped in the following way:

$$J_\alpha^{(2)} = \frac{1}{\omega}E_\beta(\mathbf{r},t)e^{-i\mathbf{q}\cdot\mathbf{r}}\int_{-\infty}^{t} dt' e^{i\omega(t-t')}\langle\psi|[j_\alpha(\mathbf{r},t), j_\beta(\mathbf{q},t')]|\psi\rangle$$

A factor of $\exp[i(\mathbf{q}\cdot\mathbf{r} - \omega t)]$ was taken outside of the integral and, to compensate, the inverse of this factor was left inside the integral. Comparing this result with (3.350) shows that they have the same form. The current $J_\alpha^{(2)}$ is proportional to the electric field. The constant of proportionality is the conductivity:

$$\sigma_{\alpha\beta}(\mathbf{q},\omega) = \frac{1}{\omega}e^{-i\mathbf{q}\cdot\mathbf{r}}\int_{-\infty}^{t} dt' e^{i\omega(t-t')}\langle\psi|[j_\alpha(\mathbf{r},t), j_\beta(\mathbf{q},t')]|\psi\rangle$$
$$+ i\frac{n_0 e^2}{m\omega}\delta_{\alpha\beta} \tag{3.383}$$

However, this result is not quite right. The final step is to average over the space variable $\mathbf{r}$ in order to eliminate atomic fluctuations. Take this average by integrating over all volume d^3r and then divide by ν. The only r dependence of the expression are the factors

$$\int d^3r e^{-i\mathbf{q}\cdot\mathbf{r}} j_\alpha(\mathbf{r},t) = j_\alpha(-\mathbf{q},t) = j_\alpha^\dagger(\mathbf{q},t) \tag{3.384}$$

The final result for the conductivity is

$$\sigma_{\alpha\beta}(\mathbf{q},\omega) = \frac{1}{\omega\nu}\int_{-\infty}^{t} dt' e^{i\omega(t-t')}\langle\psi|[j_\alpha^\dagger(\mathbf{q},t), j_\beta(\mathbf{q},t')]|\psi\rangle + i\frac{n_0 e^2}{m\omega}\delta_{\alpha\beta} \tag{3.385}$$

Equation (3.385) is the Kubo formula.

The correlation function in (3.385) is only a function of the time difference $t - t'$. If this difference is made the variable of time integration $t - t' \to t$, the formula can be expressed as

$$\sigma_{\alpha\beta}(\mathbf{q}, \omega) = \frac{1}{\omega \mathrm{v}} \int_0^\infty dt e^{i\omega t} \langle \psi | [j_\alpha^\dagger(\mathbf{q}, t), j_\beta(\mathbf{q}, 0)] | \psi \rangle + i \frac{n_0 e^2}{m\omega} \delta_{\alpha\beta}$$

which is the way it is usually written. The right-hand side is definitely not a function of t.

The wave function $|\psi\rangle$ in (3.385) is the ground state of the many-body Hamiltonian H. This Hamiltonian contains all the possible interactions in the solid except the interaction with the vector potential H'. The conductivity is calculated using (3.385), which has no mention of photon field. The conductivity is an intrinsic property of the ground state of the system. Equation (3.350) can be viewed as a Taylor series in the applied electric field $\Xi_\beta^{(\mathrm{ext})}$

$$J_\alpha(\Xi_\beta) = J_\alpha(0) + \left(\frac{\partial J_\alpha}{\partial \Xi_\beta} \right) \Xi_\beta^{(\mathrm{ext})} + O(\Xi_\beta^{(\mathrm{ext})})^2 \tag{3.386}$$

The conductivity $\sigma_{\alpha\beta} = (\partial J_\alpha / \partial \Xi_\beta)$ is calculated for $\Xi_\beta^{(\mathrm{ext})} = 0$. It is a characteristic of all linear response correlation functions that they are ground state properties. A major difficulty is that the ground state ψ is not known for most many-body systems. Then one must evaluate the correlation function using the S-matrix expansion and Green's function analysis, which has been described in Chapter 2 and in this chapter. Equation (3.385) is often the starting point for a many-body calculation, as shall be demonstrated in later chapters.

The Kubo formulas contain a retarded, two-particle correlation function. From the definition of the retarded Green's function in (3.88), the retarded correlation function of the current operator is defined as

$$\Pi_{\alpha\beta}(\mathbf{q}, t - t') = -\frac{i}{\mathrm{v}} \Theta(t - t') \langle \psi | [j_\alpha^\dagger(\mathbf{q}, t), j_\beta(\mathbf{q}, t')] | \psi \rangle \tag{3.387}$$

Its Fourier transform is

$$\Pi_{\alpha\beta}(\mathbf{q}, \omega) = -\frac{i}{\mathrm{v}} \int_{-\infty}^{\infty} dt e^{i\omega(t-t')} \Theta(t - t') \langle \psi | [j_\alpha^\dagger(\mathbf{q}, t), j_\beta(\mathbf{q}, t')] | \psi \rangle$$

Comparing this definition with (3.338), we find that

$$\sigma_{\alpha\beta}(\mathbf{q}, \omega) = \frac{i}{\omega} \left[\Pi_{\alpha\beta}(\mathbf{q}, \omega) + \frac{n_0 e^2}{m} \delta_{\alpha\beta} \right] \tag{3.388}$$

The conductivity is the retarded correlation function of the current multiplied by i and divided by ω. The correlation function $\Pi_{\alpha\beta}(\mathbf{q}, \omega)$ is usually called the *current–current correlation function*.

It is usually easiest to calculate the retarded correlation function in the Matsubara formalism. The procedure follows the prescription outlined in (3.105) and (3.110). First define the equivalent current–current correlation function in the Matsubara formalism:

$$\Pi_{\alpha\beta}(\mathbf{q}, \tau) = -\frac{1}{\mathrm{v}} \langle T_\tau j_\alpha^\dagger(\mathbf{q}, \tau) j_\beta(\mathbf{q}, 0) \rangle \tag{3.389}$$

$$\Pi_{\alpha\beta}(\mathbf{q}, i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} \Pi_{\alpha\beta}(\mathbf{q}, \tau) \tag{3.390}$$

The Matsubara function is evaluated as best one can, perhaps by using the diagrammatic techniques described in Sec. 3.4. Then the desired retarded function is obtained from

$$\underset{i\omega_n \to \omega + i\delta}{\text{change}} \Pi_{\alpha\beta}(\mathbf{q}, i\omega_n) \to \Pi_{\alpha\beta}(\mathbf{q}, \omega) \tag{3.391}$$

The conductivity is found from (3.388). These steps are always used in calculating the conductivity.

The dc conductivity is obtained by taking the limit $\mathbf{q} \to 0$ and then the limit $\omega \to 0$. The wrong answer may be obtained if the order of these limits is reversed, which may be understood on physical grounds. The limit $\omega = 0, \mathbf{q} \neq 0$ describes a static electric field, which is periodic in space. Here the charge will seek a new equilibrium, after which no current will flow. It is usually important to first take the limit of $\mathbf{q} \to 0$:

$$\lim_{\mathbf{q}\to 0} \begin{cases} \sigma_{\alpha\beta}(\mathbf{q}, \omega) &= \sigma_{\alpha\beta}(\omega) \\ \Pi_{\alpha\beta}(\mathbf{q}, i\omega_n) &= \Pi_{\alpha\beta}(i\omega_n) \\ \Pi_{\alpha\beta}(\mathbf{q}, \omega) &= \Pi_{\alpha\beta}(\omega) \\ j_\alpha(\mathbf{q}, \tau) &= j_\alpha(\tau) \end{cases} \tag{3.392}$$

The limit $\mathbf{q} \to 0$ presents no problem. The current operator is well behaved in this limit, so that the correlation functions are well behaved. In fact, when only the dc conductivity is needed, it simplifies the derivation to set $\mathbf{q} = 0$ at the beginning of the calculation.

The limit of $\omega \to 0$ is more delicate. Here the conductivity is real:

$$\mathrm{Re}\sigma_{\alpha\beta} = -\lim_{\omega\to 0} \frac{1}{\omega} \mathrm{Im}[\Pi_{\alpha\beta}(\omega)] \tag{3.393}$$

The right-hand side contains the imaginary part of the retarded correlation function. The combination $-2\mathrm{Im}[\Pi_{\alpha\beta}(\omega)]$ is just the spectral function of that operator; call it $R_{\alpha\beta}(\omega)$. By using the general formula (3.117) for the spectral function, a formal solution for $\mathrm{Im}\Pi$ is

$$\begin{aligned} -\mathrm{Im}[\Pi_{\alpha\beta}(\omega)] &= \tfrac{1}{2} R_{\alpha\beta}(\omega) \\ &= \frac{\pi}{\mathrm{V}}(1 - e^{-\beta\omega}) e^{\beta\Omega} \sum_{nm} e^{-\beta E_N} \langle n | j_a^\dagger | m \rangle \langle m | j_\beta | n \rangle \\ &\quad \times \delta(\omega + E_n - E_m) \end{aligned} \tag{3.394}$$

Now it is straightforward to take the limit $\omega \to 0$, since the prefactor of the above equation is

$$\lim_{\omega\to 0} \frac{1}{\omega}(1 - e^{-\beta\omega}) = \beta \tag{3.395}$$

and

$$\mathrm{Re}(\sigma_{\alpha\beta}) = \frac{\pi\beta}{\mathrm{V}} e^{\beta\Omega} \sum_{nm} e^{-\beta E_N} \langle n | j_a^\dagger | m \rangle \langle m | j_\beta | n \rangle \delta(E_n - E_m) \tag{3.396}$$

The conductivity is finite as $\omega \to 0$. There is no divergence in this limit, even though there is the ω^{-1} factor in (3.393).

It is a curious feature of the Kubo formula that, in order to calculate the dc conductivity, it is necessary to calculate the ac conductivity and then to take the limit $\omega \to 0$. At least this is the easiest way to do it. Later another Kubo formula will be derived strictly at $\omega = 0$. However, the latter formula is cumbersome to evaluate, and it is faster to evaluate $\sigma(\omega = 0)$ by the procedure of taking the ac conductivity and the limit that $\omega \to 0$. It is also easier, at

zero temperature, to use the Matsubara formalism and then to take the limit of $T \to 0$ at the end. Below it is shown that (3.385) is the right Kubo formula at nonzero temperatures.

The current-current correlation function is a *two-particle correlation function*. The current operator contains a product of one creation and one destruction operator, so the correlation function (3.390) contains at least four such operators:

$$\Pi_{\alpha\beta}(\mathbf{q}, \tau) = -\frac{1}{V} \sum_{\lambda\delta,\mu\nu} p_\alpha^{(\lambda\delta)}(\mathbf{q}) p_\beta^{(\mu\nu)}(\mathbf{q}) \langle T_\tau C_\lambda^\dagger(\tau) C_\delta(\tau) C_\mu^\dagger(0) C_\nu(0)\rangle$$

The correlation function describes how two particles are created and destroyed. The conductivity arises from correlations between these two events.

The conductivity is a measurable quantity. It is the same quantity as measured in a circuit or by an optical probe. Measurable quantities always involve retarded correlation functions of at least two particles. The one-particle Green's function can never be measured—at least not in the rigorous sense. The one-particle Green's function describes a Gedanken experiment which can never happen in practice. One creates a particle at time τ',

$$\mathcal{G}(\mathbf{p}, \tau - \tau') = -\langle T_\tau C_{\mathbf{p}\sigma}(\tau) C_{\mathbf{p}\sigma}^\dagger(\tau')\rangle \tag{3.397}$$

in state $\mathbf{p}\sigma$ and then destroys it at another time τ. One asks for the correlation between these events. This experiment is impossible, since real particles cannot be created or destroyed. In a two-particle Green's function, one describes a sequence of events which is realistic. A particle

$$\langle T_\tau C_\nu^\dagger(\tau) C_\mu(\tau) C_\mu^\dagger(0) C_\nu(0)\rangle \tag{3.398}$$

is changed from one state to another, say from ν to μ, and then at a later time it is changed back. The two-particle correlation function, wherein a particle changes its state, is always found in the correlation functions of linear response. In elementary particle physics, a particle can be absolutely destroyed or created. But this event always happens in conjunction with some other event which involves other particles. For example, an electron and positron can mutually annihilate and make several photons. Then one would have terms in the current operator involving the creation or destruction of two particles:

$$j_\alpha(\mathbf{q}) = \sum [p_\alpha^{(\lambda\delta)} C_\lambda d_\delta + \text{h.c.}] \tag{3.399}$$

Even in this case, the current–current correlation function involves four operators: two for one particle and two for another. If the particles interact, i.e., by a Coulomb interaction, then the correlation function may not be divided into two independent Green's functions:

$$\langle T_\tau C_\lambda(\tau) d_\delta(\tau) d_\delta^\dagger(0) C_\lambda^\dagger(0)\rangle \neq \langle T_\tau C_\lambda(\tau) C_\lambda^\dagger(0)\rangle \langle T_\tau d_\delta(\tau) d_\delta^\dagger(0)\rangle \tag{3.400}$$

We shall encounter current operators of the form (3.399). First, positron annihilation is an important experiment in metal physics (Brandt and Dupasquier, 1981; Mills *et al.*, 1986). The form of (3.399) is also used in other contexts. If an electron is destroyed in a filled band, the resulting excitation is called a *hole*. This terminology is used in the filled valence band of a semiconductor and also in the core levels of all solids. There is an identity

$$C_\lambda \left| \begin{matrix}\text{filled}\\ \text{electron}\\ \text{band}\end{matrix} \right\rangle = d_\lambda^\dagger \left| \begin{matrix}\text{empty}\\ \text{hole}\\ \text{band}\end{matrix} \right\rangle \tag{3.401}$$

where $d_\lambda^\dagger$ is the creation operator for the hole. When taking an electron from a filled band and moving it to an empty or partially filled band, then one has a current operator of the form (3.399), which is used to describe these electron–hole excitation processes.

3.8.2. Nonzero Temperatures

The preceding derivation of the Kubo formula has several restrictions. It was limited to zero temperature and transverse electric fields. Both of these restrictions are unnecessary, since (3.385) is correct for both nonzero temperature and longitudinal electric fields. At nonzero temperature, the bracket is interpreted as a thermodynamic average. The restriction to transverse electric fields was actually unnecessary. One could use a gauge wherein the scalar potential ϕ was zero but not the longitudinal vector potential. Then the longitudinal electric field can be expressed in terms of the longitudinal vector potential. The derivation is the same, step by step, with the same answer.

Another derivation is presented which is valid at nonzero temperature. A time-varying perturbation $H'(t)$ is put into the system at nonzero temperature. The central question concerns the degree to which the thermodynamic averaging is influenced by the time-varying interaction: is H' put into the thermodynamic weighting factor $\exp[\beta(\Omega - H - H' + \nu N)]$ as well as into the time development of the operators? As an example, is the following equation a correct expression for the time development of the current operator?

$$J^{(2)} \stackrel{?}{=} \mathrm{Tr}\{e^{\beta(\Omega-H-H'(t)+\mu N)}e^{it[H+H'(t)]}je^{-it[H+H'(t)]}\} \tag{3.402}$$

Although this expression is reasonable, it is believed to be wrong. It is incorrect to just put H' into the thermodynamic weighting factor. If the time oscillation is fast enough, the heat bath will not follow the oscillations. The oscillatory field causes Joule heating, but that contribution is proportional to the square of the electric field, and is not part of the linear response.

The above formula may be manipulated by moving the right-hand operator in the trace to the left, and by expressing the time operators as a time-dependent density operator:

$$J^{(2)} = \mathrm{Tr}[\rho(t)j] \tag{3.403}$$

$$\rho(t) = e^{-it[H+H'(t)]}e^{\beta(\Omega-H-H'(t)+\mu N)}e^{it[H+H'(t)]} \tag{3.404}$$

However, since the middle exponential factor also contains the operator combination $H + H'$, it commutes with the time operators, so that the above predicts

$$\rho(t) \stackrel{?}{=} e^{\beta(\Omega-H-H'(t)+\mu N)} \tag{3.405}$$

The time dependence of the current operator plays no role if the above expression is correct. It was derived by assuming H' goes into the thermodynamic weighting factors. Since our intuition is that the time development of the current operator should not cancel, the above ansatz is incorrect.

Kubo's derivation of the time-dependent density matrix proceeds in the following fashion. The density matrix $\rho_0 = \exp[\beta(\Omega - H + \mu N)]$ applies to the equilibrium system in the absence of $H'(t)$. It is assumed that the system is described by this density matrix at the initial point in time, which is taken at $t = -\infty$. The perturbation H' is adiabatically switched on as the system is brought forward in time to the present. Now the time-dependent density

matrix, for the interacting system including $H'(t)$, is called $\rho(t)$. It obeys a Heisenberg equation of motion, which when solved yields an expression for the electrical current:

$$\frac{d}{dt}\rho(t) = -i[H + H'(t), \rho(t)] \tag{3.406}$$

$$J_\alpha^{(2)} = \mathrm{Tr}[\rho(t) j_\alpha] \tag{3.407}$$

Kubo's starting point is tantamount to not including $H'(t)$ in the thermodynamic weighting factor. The density matrix is defined as $\rho(t) = \rho_0 + f(t)$, where ρ_0 is the density matrix in the absence of H'. The equilibrium density matrix ρ_0 is time independent so that the above equations simplify to

$$i\frac{d}{dt} f = [H, \rho_0] + [H, f] + [H', \rho_0] + [H', f] \tag{3.408}$$

$$[H, \rho_0] = 0 \tag{3.409}$$

The objective is to solve for the term in $J^{(2)}$, which is proportional to H', which is treated as infinitesimal. Since f is proportional to H', it follows that f is small. Terms proportional to $O(H')^2$, such as $[H', f]$, are neglected.

$$i\frac{d}{dt} f = [H, f] + [H', \rho_0] \tag{3.410}$$

This equation is solved by moving the first term on the right to the left:

$$i\frac{d}{dt} f - [H, f] = [H', \rho_0] \tag{3.411}$$

and then expressing the left-hand side as

$$e^{-itH}\left\{i\frac{d}{dt}(e^{itH} f e^{-itH})\right\}e^{itH} = i\frac{d}{dt} f - [H, f] = [H', \rho_0] \tag{3.412}$$

The linear differential equation may be integrated to give

$$i\frac{d}{dt}(e^{itH} f e^{-itH}) = e^{itH}[H', \rho_0]e^{-itH} = [H'(t), \rho_0] \tag{3.413}$$

$$f(t) = f(-\infty) - ie^{-itH}\left\{\int_{-\infty}^{t} dt'[H'(t'), \rho_0]\right\}e^{itH} \tag{3.414}$$

The interaction is switched on slowly in time, so that at $t = -\infty$ there is no interaction, and $H' = 0$. Of course, then $f(-\infty) = 0$, since this term only exists when H' exists. A solution for f has been derived which is proportional to H', which is adequate for a description of linear response.

The evaluation of (3.407) is

$$J_\alpha^{(2)} = \mathrm{Tr}[\rho_0 j_\alpha] + \mathrm{Tr}[f(t) j_\alpha] \tag{3.415}$$

where the first term on the right vanishes ($\mathrm{Tr}[\rho_0 j_\alpha] = 0$) since there is no current if there is no external applied field. The expectation value of the current, proportional to H', is the remaining term:

$$J_\alpha^{(2)}(\mathbf{r}, t) = -i\mathrm{Tr}\left\{e^{-itH}\left[\int_{-\infty}^{t} dt'[H'(t'), \rho_0]\right]e^{itH} j_\alpha(\mathbf{r})\right\} \tag{3.416}$$

$$= -i\mathrm{Tr}\left\{\int_{-\infty}^{t} dt'[H'(t'), \rho_0]e^{itH} j_\alpha(\mathbf{r})e^{-itH}\right\} \tag{3.417}$$

$$= -i\mathrm{Tr}\left\{\int_{-\infty}^{t} dt'[H'(t'), \rho_0] j_\alpha(\mathbf{r}, t)\right\} \tag{3.418}$$

The quantity in curly braces was rearranged by using the cyclic properties of the trace. The factors on the right are just the time development of the current operator: by using the cyclic properties of the trace, the three operators can be rearranged:

$$\begin{aligned}\mathrm{Tr}[H'(t'), \rho_0] j_\alpha(\mathbf{r}, t) &= \mathrm{Tr}[H'(t')\rho_0 j_\alpha(\mathbf{r}, t) - \rho_0 H'(t') j_\alpha(\mathbf{r}, t)] \\ &= \mathrm{Tr}\{\rho_0[j_\alpha(\mathbf{r}, t)H'(t') - H'(t') j_\alpha(\mathbf{r}, t)]\} \\ &= \mathrm{Tr}\{\rho_0[j_\alpha(\mathbf{r}, t), H'(t')]\}\end{aligned} \tag{3.419}$$

The term in curly braces is just a commutator. The final result is derived for the expectation value of the current, which is just the thermodynamic average of the commutator of $H'(t')$ and $j_\alpha(\mathbf{r}, t)$:

$$J_\alpha^{(2)}(\mathbf{r}, t) = -i\int_{-\infty}^{t} dt' \langle[j_\alpha(\mathbf{r}, t), H'(t')]\rangle \tag{3.420}$$

Equation (3.420) is precisely the zero-temperature equation which was derived earlier in (3.381). It leads, in the same way, to the Kubo formula (3.385). Previously the angle brackets meant to take the expectation value in the ground state at zero temperature, while now they mean a trace over the thermal distribution at nonzero temperatures. Equation (3.385) is valid at nonzero temperatures.

3.8.3. Zero Frequency

The Kubo formula (3.385) has been shown to be valid at nonzero temperatures and for transverse electric fields. Also, it was suggested that it is valid for longitudinal electric fields. However, the purist might still object. All the derivations introduce a frequency ω, and the dc conductivity is obtained by letting $\omega \to 0$ at the end of the derivation. Is it possible to provide a derivation of the dc conductivity which assumes a constant electric field throughout the derivation? Of course, the answer is affirmative, which provides the motivation for the next derivation, using a method suggested by Luttinger (1964).

Begin by assuming an electric field $\mathbf{E}_0$ which is static in time and constant in space. It causes a scalar potential

$$\phi(\mathbf{r}) = -\mathbf{E}_0 \cdot \mathbf{r} \tag{3.421}$$

and the external vector potential is zero. This scalar potential introduces a perturbation term into the Hamiltonian which is

$$e^{st}F = e^{st}\int \rho(\mathbf{r})\phi(\mathbf{r}) \tag{3.422}$$

where $\rho(\mathbf{r})$ is the charge density operator of the system. A factor of $\exp(st)$ was added to this term so that it vanishes as $t \to -\infty$. This factor represents the switching on of the potential as time develops from the past. It is assumed that the perturbation on the density matrix,

$$\rho = \rho_0 + fe^{st} \tag{3.423}$$

also has this term. The first term $\rho_0 = \exp[\beta(\Omega - H)]$ is the density matrix when F is absent. The perturbation fe^{st} is due to F. The starting point is the equation of motion for the density operator,

$$i\frac{\partial\rho}{\partial t} = [H + e^{st}F, \rho + fe^{st}] \tag{3.424}$$

which is solved as before. The term proportional to the perturbation F is given in (3.414):

$$fe^{st} = -i\int_{-\infty}^{t} dt' e^{-iHt}e^{st'}[F(t'), \rho_0]e^{itH} = -i\int_{-\infty}^{t} dt' e^{st'}[F(t'-t), \rho_0]$$

It is convenient to change integration variables to $t' - t \Rightarrow -t$:

$$f = -i\int_0^{\infty} dt e^{-st}[F(-t), \rho_0] \tag{3.425}$$

The crucial trick is to write the following identity for the commutator:

$$[F(-t), \rho_0] = -i\rho_0\int_0^{\beta} d\beta' \frac{\partial}{\partial t}F(-t - i\beta') \tag{3.426}$$

The identity is easy to prove,

$$\begin{aligned}\int_0^{\beta} d\beta' \frac{\partial}{\partial t}F(-t - i\beta') &= -i\int_0^{\beta} d\beta' \frac{\partial}{\partial \beta'}F(-t - i\beta') \\ &= -i[F(-t - i\beta) - F(-t)]\end{aligned}$$

so that

$$\begin{aligned}i^2\rho_0[F(-t - i\beta) - F(-t)] &= -\rho_0[e^{\beta H}F(-t)e^{-\beta H} - F(-t)] \\ &= -[F(-t), \rho_0]\end{aligned} \tag{3.427}$$

The expectation value of the current is the expression

$$J_\alpha(\mathbf{r}) = \mathrm{Tr}[fj_\alpha(\mathbf{r})] = -\int_0^{\infty} dt e^{-st}\int_0^{\beta} d\beta' \mathrm{Tr}\left[\rho_0 \frac{\partial F(-t - i\beta')}{\partial t} j_\alpha(\mathbf{r})\right] \tag{3.428}$$

The reason that the trick identity is introduced is that one can now do the following manipulations to produce the electric field. First, the time development of $F(-t)$ is given by

$$F(-t) = \int d^3r\rho(\mathbf{r}, -t)\phi(\mathbf{r}) \tag{3.429}$$

The scalar potential is time independent. The time derivative of the density is related to the current through the equation of continuity:

$$\frac{\partial F(-t)}{\partial t} = \int d^3r\phi(\mathbf{r})\frac{\partial\rho(\mathbf{r}, -t)}{\partial t} \tag{3.430}$$

$$= -\int d^3r\phi(\mathbf{r})\nabla\cdot\mathbf{j}(\mathbf{r}, -t) \tag{3.431}$$

Integrate by parts on the space variable, and thereby obtain the electric field $\mathbf{E}_0$:

$$\frac{\partial F(-t)}{\partial t} = \int d^3r\mathbf{j}(\mathbf{r}, -t)\cdot\nabla\phi(\mathbf{r}) \tag{3.432}$$

$$= -\int d^3r\mathbf{j}(\mathbf{r}, -t)\cdot\mathbf{E}_0 \tag{3.433}$$

$$J_\alpha(\mathbf{r}) = \int_0^\infty dte^{-st}\int_0^\beta d\beta'\mathrm{Tr}\left[\rho_0\int d^3r'j\beta(\mathbf{r}', -t - i\beta')E_{0\beta}\,j\alpha(\mathbf{r})\right]$$

Since the electric field is a constant, it can be removed from the integrals, and the conductivity is given by

$$\mathrm{Re}(\sigma_{\alpha\beta}) = \int_0^\infty dte^{-st}\int_0^\beta d\beta'\mathrm{Tr}\left[\rho_0\int d^3r'j_\beta(\mathbf{r}', -t - i\beta')j_\alpha(\mathbf{r})\right]$$

The right-hand side appears to depend on the special variable $\mathbf{r}$. This dependence is eliminated, as before, by averaging over the volume ν of the sample. This manipulation just introduces the $\mathbf{q} \to 0$ limit of the current operator $j_\alpha(\mathbf{q})$

$$\lim_{\mathbf{q}\to 0} j_\alpha(\mathbf{q}, t) = j_\alpha(t) = \int d^3rj_\alpha(\mathbf{r}, t) \tag{3.434}$$

The Kubo formula for the dc electrical conductivity is

$$\mathrm{Re}(\sigma_{\alpha\beta}) = \frac{1}{\nu}\int_0^\infty dte^{-st}\int_0^\beta d\beta'\mathrm{Tr}\left[\rho_0\int d^3r'j_\beta(-t - i\beta')j_\alpha\right] \tag{3.435}$$

The result does not contain the frequency. The elimination of ω has been achieved by paying a penalty: there is a second integration, this time over the temperature-like variable β'. The second integration makes the results appear different than the prior result (3.385). The two results are identical for $\omega = 0$. To prove this assertion, again introduce the representation $|n\rangle$, $|m\rangle$, which are exact eigenstates of the Hamiltonian H. The important matrix element is

$$\langle n|j_\beta(-t - i\beta')|m\rangle = \langle n|e^{-iH(t+i\beta')}j_\beta e^{iH(t+i\beta')}|m\rangle \tag{3.436}$$

$$= \langle n|j_\beta|m\rangle e^{-it(E_n-E_m)}e^{\beta'(E_n-E_m)} \tag{3.437}$$

so that

$$\mathrm{Re}(\sigma_{\alpha\beta}) = \frac{1}{\nu}\mathrm{Re}\left[\sum_{nm} e^{-\beta E_n}\langle n|j_\beta|m\rangle\langle m|j_\alpha|n\rangle \times\int_0^\infty dte^{-it(E_n-E_m-is)}\int_0^\beta d\beta' e^{\beta'(E_n-E_m)}\right] \tag{3.438}$$

Both integrals can be done easily. For the real part of the conductivity, the only complex term is the t integral:

$$\mathrm{Re}\left[\int_0^\infty dt e^{-it(E_n-E_m-is)}\right] = \mathrm{Re}\left[\frac{1}{i(E_n-E_m-is)}\right] = \pi\delta(E_n-E_m)$$

$$\delta(E_n-E_m)\int_0^\beta d\beta' e^{\beta'(E_n-E_m)} = \beta\delta(E_n-E_m)$$

which makes the integral over β' easy. The final result is

$$\mathrm{Re}(\sigma_{\alpha\beta}) = \frac{\pi\beta}{\nu}\sum_{nm} e^{-\beta E_n}\langle n|j_\beta|m\rangle\langle m|j_\alpha|n\rangle\delta(E_n-E_m)$$

This equation is identical to the earlier result (3.396), since $j_\alpha^\dagger = j_\alpha$ for $\mathbf{q}=0$. It has been proved that (3.435) is identical to the $\omega = 0$ limit of the usual Kubo formula (3.385). In the limit of $\omega = 0$, the longitudinal and transverse conductivities are identical.

3.8.4. Photon Self-Energy

A formula which is identical to the Kubo formula can also be derived from the photon self-energy. Consider the photon Green's function (3.78),

$$\mathcal{D}_{\mu\nu}(\mathbf{k},\tau) = -\langle T_\tau A_\mu(\mathbf{k},\tau)A_\nu(-\mathbf{k},0)\rangle \tag{3.439}$$

and treat as the perturbation

$$V = -\frac{1}{c}\int d^3r\left[\mathbf{j}(\mathbf{r})\cdot\mathbf{A}(\mathbf{r}) + \frac{e^2 n(\mathbf{r})}{2mc}A(\mathbf{r})^2\right] \tag{3.440}$$

$$= -\frac{1}{\sqrt{\nu}}\sum_{\mathbf{k}}\mathbf{j}(\mathbf{k})\cdot\mathbf{A}(\mathbf{k}) + \frac{e^2 n_0}{2m}\sum_{\mathbf{k}}\mathbf{A}(\mathbf{k})\cdot\mathbf{A}(-\mathbf{k}) \tag{3.441}$$

The $A(\mathbf{k})^2$ term contributes a self-energy of $\Pi^{(1)} = n_0 e^2/m$. The other term in the self-energy comes from the $n=2$ term in the expansion for the S matrix for the $\mathbf{j}\cdot\mathbf{A}$ interaction,

$$-\frac{1}{2!\nu}\sum_{\mathbf{k}'\mathbf{k}''}\int_0^\beta d\tau_1\int_0^\beta d\tau_2\langle T_\tau j_l(\mathbf{k}',\tau_1)j_m(\mathbf{k}'',\tau_2)\rangle \tag{3.442}$$

$$\times\langle T_\tau A_\mu(\mathbf{k},\tau)A_l(\mathbf{k}',\tau_1)A_m(\mathbf{k}'',\tau_2)A_\nu(-\mathbf{k},0)\rangle$$

$$= \sum_{\mathbf{k}}\int_0^\beta d\tau_1\int_0^\beta d\tau_2\mathcal{D}^{(0)}_{\mu l}(\mathbf{k},\tau-\tau_1)\Pi_{lm}(\mathbf{k},\tau_1-\tau_2)\mathcal{D}^{(0)}_{m\nu}(\mathbf{k},\tau_2) \tag{3.443}$$

where

$$\Pi_{lm}(\mathbf{k},\tau_1-\tau_2) = -\frac{1}{\nu}\langle T_\tau j_l(-\mathbf{k},\tau_1)j_m(\mathbf{k},\tau_2)\rangle \tag{3.444}$$

$$= -\frac{1}{\nu}\langle T_\tau j_l^\dagger(\mathbf{k},\tau_1)j_m(\mathbf{k},\tau_2)\rangle \tag{3.445}$$

The Fourier transform of this expression gives the self-energy contribution as

$$\Pi_{lm}(\mathbf{k},i\omega_n) = -\frac{1}{\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau j_l^\dagger(\mathbf{k},\tau)j_m(\mathbf{k},0)\rangle \tag{3.446}$$

Recall Eq. (2.186) for the dielectric function, in retarded or Matsubara notation:

$$\varepsilon(\mathbf{k}, i\omega_n) = 1 - \frac{4\pi}{(i\omega_n)^2}\left[\Pi(\mathbf{k}, i\omega_n) + \frac{n_0 e^2}{m}\right] \tag{3.447}$$

$$\varepsilon_{\text{ret}}(\mathbf{k}, \omega) = 1 - \frac{4\pi}{(\omega)^2}\left[\Pi(\mathbf{k}, \omega) + \frac{n_0 e^2}{m}\right] \tag{3.448}$$

where Π is the scalar part of the self-energy $\Pi_{\alpha\beta}$ for homogeneous, isotopic systems:

$$\Pi(\mathbf{k}, i\omega_n) = \frac{1}{3}\sum_{\alpha} \Pi_{\alpha\alpha}(\mathbf{k}, i\omega_n) \tag{3.449}$$

The long-wavelength dielectric function is given by the conductivity

$$\varepsilon(\omega) = \lim_{\mathbf{k}\to 0} \varepsilon_{\text{ret}}(\mathbf{k}, \omega) = 1 - \frac{4\pi\sigma(\omega)}{i\omega} \tag{3.450}$$

Comparing (3.448) with (3.450), the expression for the conductivity $\sigma(\omega)$ in terms of the retarded correlation function $\Pi(\omega)$ is

$$\sigma(\omega) = \frac{i}{\omega}\left[\Pi(\omega) + \frac{n_0 e^2}{m}\right] \tag{3.451}$$

Since Π is the current–current correlation function, Eq. (3.388) is rederived.

The important point is reemphasized that, in evaluating the Kubo formula for the current–current correlation function, do not include the self-energy terms which arise from the vector potential $\mathbf{A}(\mathbf{r}, t)$. These terms have already been included in the formalism, indirectly, by changing the external field $\mathbf{E}^{(\text{ext})}$ to the total field $\mathbf{E}$. They do not reenter the formalism again except as internal interactions within the polarization bubble.

3.9. OTHER KUBO FORMULAS

There are other measurements besides electrical conductivity. They require additional correlation functions or Kubo formulas. Some of these will be derived here. The first is magnetic susceptibility. Next comes the derivation of transport coefficients such as thermal conductivity or thermoelectric power.

3.9.1. Pauli Paramagnetic Susceptibility

The magnetic susceptibility of a solid depends rather strikingly on whether the system is already spontaneously magnetized. If it is, then an additional weak, external magnetic field may not change the magnetization appreciably—there may be no linear response. This case is not considered, and it is assumed the solid is not in a magnetic state; it is not ferromagnetic, antiferromagnetic, etc. A magnetic state has long-range order, although there is usually short-range ordering, from the interactions, even in the absence of long-range order.

The perturbation term in the Hamiltonian for Pauli paramagnetism has just the interaction with the electron spin:

$$V = -\mathbf{m} \cdot \mathbf{H}_0 e^{-i\omega t} \tag{3.452}$$

$$\mathbf{H}(t) = \mathbf{H}_0 e^{-i\omega t} \tag{3.453}$$

$$m_\alpha = g\mu_0 \sum_i S_{i,\alpha} \tag{3.454}$$

where $\mathbf{H}(t)$ is the ac magnetic field and $\mathbf{m}$ is the magnetization, which is the sum over all spins $S_{i,\alpha}$ at position i with vector direction α. The other factors are the Bohr magneton $\mu_0 = |e|\hbar/2mc$ and gyromagnetic ratio $g \approx 2$. Another contribution to the magnetization is the orbital, or Landau, contribution, which is not discussed here. In Pauli paramagnetism the objective is to evaluate the magnetization,

$$M_\alpha(t) = \langle m_\alpha(t) \rangle = \chi_{\alpha\beta} H_\beta(t) \tag{3.455}$$

for terms linear in the magnetic field. The derivation proceeds exactly as in the preceding section. In analogy with (3.381), the magnetization is a correlation function

$$M_\alpha(t) = -i \int_{-\infty}^{t} dt' \langle [m_\alpha(t), V(t')] \rangle \tag{3.456}$$

and the magnetic susceptibility is

$$\chi_{\alpha\beta}(\omega) = i \int_{-\infty}^{t} dt' e^{i\omega(t-t')} \langle [m_\alpha(t), m_\beta(t')] \rangle \tag{3.457}$$

Equation (3.457) is just the Fourier transform of a retarded correlation function. It may be evaluated conveniently in the Matsubara formalism. Define a correlation function

$$\chi_{\alpha\beta}(i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau m_\alpha(\tau) m_\beta(0) \rangle \tag{3.458}$$

and (3.457) is just the retarded function obtained by the analytic continuation $i\omega_n \to \omega + i\delta$. The spin operators are often written in terms of raising and lowering operators:

$$S^{(\pm)} = S_x \pm iS_y \tag{3.459}$$

$$S^{(+)} = \sum_{\mathbf{k}} C^\dagger_{\mathbf{k}\uparrow} C_{\mathbf{k}\downarrow} \tag{3.460}$$

The three operators $S^{(+)}, S^{(-)}, S_z$ are used instead of S_x, S_y, S_z. Of course, susceptibility functions can also be constructed using these operators.

The correlation function will be evaluated for a simple example. It is the Pauli susceptibility of a free-electron gas at zero temperature. The answer is first derived using simple considerations. In a magnetic field H_0 in the z direction, the spin-up and spin-down electrons have the following energies and occupation functions

$$\xi_\uparrow(\mathbf{p}) = \xi_\mathbf{p} - \mu_0 H_0, \; n_\uparrow = n_F(\xi_\mathbf{p} - \mu_0 H_0) \tag{3.461}$$

$$\xi_\downarrow(\mathbf{p}) = \xi_\mathbf{p} + \mu_0 H_0, \; n_\downarrow = n_F(\xi_\mathbf{p} + \mu_0 H_0) \tag{3.462}$$

The net magnetization will be the difference between the number of spin-up and spin-down electrons. The occupation numbers are expanded in a power series in the magnetic field. Keeping the first two terms gives

$$M_z = \mu_0 \sum_{\mathbf{p}} [n_\uparrow(\mathbf{p}) - n_\downarrow(\mathbf{p})] = -2\mu_0^2 H_0 \sum_{\mathbf{p}} \frac{\partial}{\partial \xi_{\mathbf{p}}} n_F(\xi_{\mathbf{p}}) \tag{3.463}$$

$$\chi_p = \lim_{T\to 0} \frac{M_z}{H_0} = 2\mu_0^2 \sum_{\mathbf{p}} \delta(\xi_{\mathbf{p}}) = \mu_0^2 N_F \tag{3.464}$$

The Pauli susceptibility χ_p is given by the square of the Bohr magneton μ_0 times the density of states N_F at the Fermi energy. This density of states is the total for both spin components. The notation is confusing, since often the same symbol is used as the density of states for each spin component, so that the value is only one-half as large.

The same result is obtained from the correlation function. The operator m_α is given in terms of creation and destruction operators as

$$m_z = \mu_0 \sum_{\mathbf{p}\sigma} \sigma C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} \tag{3.465}$$

where $\sigma = \pm$ is used to denote electron spin up or down. The Kubo formula for the correlation function for the static susceptibility ($i\omega = 0$) is

$$\chi_{zz} = \mu_0^2 \int_0^\beta d\tau \sum_{\mathbf{pk},\sigma\sigma'} \sigma\sigma' \langle T_\tau C^\dagger_{\mathbf{k}\sigma}(\tau) C_{\mathbf{k}\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma'}(0) C_{\mathbf{p}\sigma'}(0)\rangle \tag{3.466}$$

For the noninteracting system, the Hamiltonian is just the kinetic energy term, so that $C^\dagger_{\mathbf{k}\sigma}(\tau) = e^{\tau\xi_{\mathbf{k}}} C^\dagger_{\mathbf{k}\sigma}$ and $C_{\mathbf{k}\sigma}(\tau) = e^{-\tau\xi_{\mathbf{k}}} C_{\mathbf{k}\sigma}$. Then the operator $C^\dagger_{\mathbf{k}\sigma}(\tau) C_{\mathbf{k}\sigma}(\tau)$ has no τ dependence, so the τ integral just gives β. The combination of operators can be averaged directly. Some care must be taken with the terms which have $\sigma = \sigma'$ and $\mathbf{p} = \mathbf{k}$. Using the fact that $\langle n_{\mathbf{k}}^2\rangle = \langle n_{\mathbf{k}}\rangle$ if both refer to the same spin state, the answer is

$$\chi_{zz} = \mu_0^2 \beta \left(\sum_{\mathbf{p}\sigma} \sigma^2 n_{\mathbf{p}} + \sum_{\mathbf{p}\neq\mathbf{k},\sigma\sigma'} \sigma\sigma' n_{\mathbf{k}} n_{\mathbf{p}} + \sum_{\mathbf{p},\sigma\neq\sigma'} \sigma\sigma' n_{\mathbf{p}}^2 \right) \tag{3.467}$$

$$= \mu_0^2 \beta \left[\sum_{\mathbf{p}\sigma} \sigma^2 (n_{\mathbf{p}} - n_{\mathbf{p}}^2) + \left(\sum_{\mathbf{p}\sigma} \sigma n_{\mathbf{p}} \right)^2 \right] \tag{3.468}$$

The second term on the right vanishes, since there are as many up spins as down spins in equilibrium. The correlation function is evaluated, of course, in the absence of any external magnetic field. There remains the combination of operators

$$\chi_{zz} = \mu_0^2 \beta \sum_{\mathbf{p}\sigma} n_{\mathbf{p}}(1 - n_{\mathbf{p}}) = -\mu_0^2 \sum_{\mathbf{p}\sigma} \left(\frac{\partial n_F}{\partial \xi_{\mathbf{p}}} \right) = 2\mu_0^2 \sum_{\mathbf{p}\sigma} \delta(\xi_{\mathbf{p}}) \tag{3.469}$$

This formula has exactly the same combination of factors as obtained previously, which confirms the result of the simple derivation.

Another method of evaluating the correlation function is to define the wave-vector-dependent susceptibility $\chi(\mathbf{q}, i\omega)$ by

$$\chi_{\alpha\beta}(\mathbf{q}, i\omega) = \int_0^\beta d\tau e^{i\omega_n\tau} \langle T_\tau m_\alpha(\mathbf{q}, \tau) m_\beta(-\mathbf{q}, 0)\rangle \tag{3.470}$$

$$m_z(\mathbf{q}) = \mu_0 \sum_{\mathbf{p}\sigma} \sigma C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{3.471}$$

$$m_x(\mathbf{q}) = \frac{\mu_0}{\sqrt{2}} \sum_{\mathbf{p}} (C^\dagger_{\mathbf{p}+\mathbf{q}\uparrow} C_{\mathbf{p}\downarrow} + C^\dagger_{\mathbf{p}+\mathbf{q}\downarrow} C_{\mathbf{p}\uparrow}) \tag{3.472}$$

$$m_y(\mathbf{q}) = \frac{i\mu_0}{\sqrt{2}} \sum_{\mathbf{p}} (C^\dagger_{\mathbf{p}+\mathbf{q}\uparrow} C_{\mathbf{p}\downarrow} - C^\dagger_{\mathbf{p}+\mathbf{q}\downarrow} C_{\mathbf{p}\uparrow}) \tag{3.473}$$

Evaluate χ_{zz} for nonzero values of $\mathbf{q}$ and $i\omega = 0$ and formally take the limit $\mathbf{q} \to 0$ at the end of the calculation:

$$\chi_{zz}(\mathbf{q}, 0) = \mu_0^2 \sum_{\mathbf{pk},ss'} ss' \int_0^\beta d\tau \langle T_\tau C^\dagger_{\mathbf{p}+\mathbf{q},s}(\tau) C_{\mathbf{p},s}(\tau) C^\dagger_{\mathbf{k}-\mathbf{q},s'}(0) C_{\mathbf{k},s'}(0)\rangle$$

For noninteracting electrons, this correlation function is nonzero only if $s = s'$, and then it is identical to the polarization operator $P^{(1)}(\mathbf{q}, 0)$ which has been evaluated previously:

$$\chi_{zz}(\mathbf{q}, 0) = -\mu_0^2 P^{(1)}(\mathbf{q}, 0) = -2\mu_0^2 \sum_{\mathbf{p}} \frac{n_{\mathbf{p}} - n_{\mathbf{p}+\mathbf{q}}}{\xi_{\mathbf{p}} - \xi_{\mathbf{p}+\mathbf{q}}} \tag{3.474}$$

Now take the limit of $\mathbf{q} \to 0$ on the right-hand side. There results $\partial n_F(\xi_{\mathbf{p}})/\partial \xi_{\mathbf{p}}$ in the summation over $\mathbf{p}$. All methods give the same result for the noninteracting electron system.

3.9.2. Thermal Currents and Onsager Relations

There are two conceptual difficulties in deriving a formula for the thermal conductivity. The first is that one must put a temperature gradient on the solid. However, all of our thermal averaging assumes a constant temperature. Of course, the thermal conductivity is still defined in the limit that the temperature difference ΔT goes to zero, so that the correlation function can be evaluated at a single temperature. But the problem lies in formally deriving the correlation function without assuming that ΔT exists somewhere in the formalism. Another way to express this difficulty is to consider how one raises the temperature by heating the sample. The heating changes the energy of the system, so that one might try to calculate the time rate of change of the energy. But the energy operator is the Hamiltonian, and its rate of change is $\dot{H} = i[H, H] = 0$, which does not help much. This problem has been well studied in nonequilibrium statistical mechanics (de Groot, 1952, 1962), and it has been found that the important quantity is the rate of change of the entropy.

Another problem with the thermal conductivity is that there are several definitions of heat current. We shall mention three. The same definition of experimental quantities is obtained for the thermal conductivity, thermo-electrical power, etc., regardless of which definition is adopted. But one must be careful to do the right calculation for each definition of heat current.

In linear response, there are current $\mathbf{J}_i$ which flow as a result of "forces" $\mathbf{X}_i$ on the system. These forces might be temperature gradients ∇T, or electric fields $\mathbf{E} = -\nabla V$, or

concentration gradients which are expressed as gradients of the chemical potential $\nabla\mu$. Linear response assumes these are proportional:

$$\mathbf{J}_i = \sum_j \mathbf{Z}^{(ij)} \cdot \mathbf{X}_j \tag{3.475}$$

The coefficients $\mathbf{Z}^{(ij)}$ are the measurable constants which are defined in terms of correlation functions. There are also Onsager relationships, which specify that $\mathbf{Z}^{(ij)} = \mathbf{Z}^{(ji)}$. A moment's reflection will show that the Onsager relationship is not valid for any arbitrary choice of currents and forces. For example, if they are valid for a force $\nabla(1/T) = -(\nabla T)/T^2$, then changing the force to ∇T, so that $1/T^2$ is absorbed in $\mathbf{Z}$, means the Onsager relation is no longer valid. There must be a criterion for choosing the forces and currents, which has been given in nonequilibrium statistical mechanics by de Groot (1952). In a nonequilibrium process, there is a net generation of entropy, so that $\partial S/\partial t > 0$. Here S is just the part of the entropy which is generated irreversibly. By requiring that entropy generation be expressed as

$$\frac{\partial S}{\partial t} = \sum_i \mathbf{J}_i \cdot \mathbf{X}_i \tag{3.476}$$

then the Onsager relations are valid. There is not a unique set of currents and forces, and many possibilities satisfy (3.476). Each choice defines a different set of coefficients $\mathbf{Z}^{(ij)}$, but each set obeys the relationship $\mathbf{Z}^{(ij)} = \mathbf{Z}^{(ji)}$.

The symbol $\mathbf{J}$ will refer to particle current. There are several definitions of energy current. The discussion will follow Barnard (1972). The first energy current which comes to mind is the one called the energy current $\mathbf{J}_E$. For a free-particle system it is just the velocity of the particles times the energy times the number:

$$\mathbf{j}_E = \sum_{\mathbf{p}\sigma} \mathbf{v}_\mathbf{p} \varepsilon_\mathbf{p} n_\mathbf{p} \tag{3.477}$$

$$\mathbf{J}_E = \langle \mathbf{j}_E \rangle \tag{3.478}$$

Other definitions were given in Sec. 1.2 for particles which were not free. For a system with only these two currents, their forces are (Luttinger, 1964)

$$\begin{aligned} \mathbf{J}_1 &= \mathbf{J}, \qquad \mathbf{X}_1 = -\frac{e}{T}\nabla V - \nabla\left(\frac{\mu}{T}\right) \\ \mathbf{J}_2 &= \mathbf{J}_E, \quad \mathbf{X}_2 = \nabla\left(\frac{1}{T}\right) \end{aligned} \tag{3.479}$$

where e is the charge of the particle. The rate of entropy production is

$$\frac{\partial S}{\partial t} = -\mathbf{J} \cdot \left[\frac{e}{T}\nabla V + \nabla\left(\frac{\mu}{T}\right)\right] + \mathbf{J}_E \cdot \nabla\left(\frac{1}{T}\right) \tag{3.480}$$

and the linear response equations (3.475) are

$$J_\alpha = -M_{\alpha\beta}^{(11)}\left[\frac{e}{T}\nabla_\beta V + \nabla_\beta\left(\frac{\mu}{T}\right)\right] + M_{\alpha\beta}^{(12)}\nabla_\beta\left(\frac{1}{T}\right) \tag{3.481}$$

$$J_{E,\alpha} = -M_{\alpha\beta}^{(21)}\left[\frac{e}{T}\nabla_\beta V + \nabla_\beta\left(\frac{\mu}{T}\right)\right] + M_{\alpha\beta}^{(22)}\nabla_\beta\left(\frac{1}{T}\right) \tag{3.482}$$

With these definitions, the Onsager relation is $M_{\alpha\beta}^{(12)} = M_{\alpha\beta}^{(21)}$.

The trouble with (3.477) is that the energy current is not the heat current. As discussed by Taylor (1970), the heat current should be

$$\mathbf{J}_Q = \mathbf{J}_E - \mu\mathbf{J} \tag{3.483}$$

which for free particles has the form

$$\mathbf{j}_Q = \sum_{\mathbf{p}\sigma} \mathbf{v}_{\mathbf{p}} n_{\mathbf{p}} (\varepsilon_{\mathbf{p}} - \mu) \tag{3.484}$$

so that the reference energy is the chemical potential μ. The latter definition makes sense only for particles with a positive chemical potential, e.g., electrons in metals. If one takes an electron below the chemical potential and moves it down to the other end of the sample, one has moved energy. But when the electron arrives at the new location, it finds locally a filled Fermi distribution, so that the only states available to it are above the Fermi energy. But the energy gain required to increase its energy, to get it above the Fermi energy, must come from the surroundings. The energy gain from the surroundings must cool the locality, so that the electron has brought coldness with it. Only electrons which arrive with energy above the Fermi energy, and give energy to the surroundings, bring heat.

With this choice of thermal current, the forces are

$$\begin{aligned} \mathbf{J}_1 &= \mathbf{J}, \qquad \mathbf{X}_1 = -\frac{1}{T}\nabla\bar{\mu} \\ \mathbf{J}_2 &= \mathbf{J}_Q, \quad \mathbf{X}_2 = \nabla\left(\frac{1}{T}\right) \end{aligned} \tag{3.485}$$

where $\bar{\mu} = \mu + eV$. With this choice, the rate of entropy production is

$$\frac{\partial S}{\partial t} = -\frac{1}{T}\mathbf{J}\cdot\nabla\bar{\mu} + \mathbf{J}_Q\cdot\nabla\left(\frac{1}{T}\right) \tag{3.486}$$

and the linear response equations are

$$J_\alpha = -\frac{1}{T}L^{(11)}_{\alpha\beta}\nabla_\beta\bar{\mu} + L^{(12)}_{\alpha\beta}\nabla_\beta\left(\frac{1}{T}\right) \tag{3.487}$$

$$J_{Q\alpha} = -\frac{1}{T}L^{(21)}_{\alpha\beta}\nabla_\beta\bar{\mu} + L^{(22)}_{\alpha\beta}\nabla_\beta\left(\frac{1}{T}\right) \tag{3.488}$$

where $L^{(12)} = L^{(21)}$. It is easy to show that these equations are consistent with the set (3.479) and (3.481). In (3.479) regroup the terms,

$$\frac{e}{T}\nabla V + \nabla\left(\frac{\mu}{T}\right) = \frac{e}{T}\nabla V + \frac{1}{T}\nabla\mu + \mu\nabla\left(\frac{1}{T}\right) = \frac{1}{T}\nabla\bar{\mu} + \mu\nabla\left(\frac{1}{T}\right)$$

so that $\partial S/\partial t$ becomes

$$\frac{\partial S}{\partial t} = -\mathbf{J}\cdot\left[\frac{1}{T}\nabla\bar{\mu} + \mu\nabla\left(\frac{1}{T}\right)\right] + \mathbf{J}_E\cdot\nabla\left(\frac{1}{T}\right) \tag{3.489}$$

$$= -\frac{1}{T}\mathbf{J}\cdot\nabla\bar{\mu} + (\mathbf{J}_E - \mu\mathbf{J})\cdot\nabla\left(\frac{1}{T}\right) \tag{3.490}$$

which is identical to (3.486). Similarly, in order to find the relationship between the coefficients $M^{(ij)}$, and $L^{(ij)}$, operate on the current equations (3.481) with the same regrouping

$$J_\alpha = -\frac{1}{T}M^{(11)}_{\alpha\beta}\nabla_\beta\bar{\mu} + (M^{(12)}_{\alpha\beta} - \mu M^{(11)}_{\alpha\beta})\nabla_\beta\left(\frac{1}{T}\right) \tag{3.491}$$

$$J_{E\alpha} = -\frac{1}{T}M^{(21)}_{\alpha\beta}\nabla_\beta\bar{\mu} + (M^{(22)}_{\alpha\beta} - \mu M^{(21)}_{\alpha\beta})\nabla_\beta\left(\frac{1}{T}\right) \tag{3.492}$$

Next, consider the following combination of these equations:

$$\begin{aligned} J_{Q\alpha} = J_{E,\alpha} - \mu J_\alpha &= -\frac{1}{T}(M^{(21)}_{\alpha\beta} - \mu M^{(11)}_{\alpha\beta})\nabla_\beta\bar{\mu} \\ &+ (M^{(22)}_{\alpha\beta} - \mu M^{(21)}_{\alpha\beta} - \mu M^{(12)}_{\alpha\beta} + \mu^2 M^{(11)}_{\alpha\beta})\nabla_\beta\left(\frac{1}{T}\right) \end{aligned} \tag{3.493}$$

This equation for $J_{Q,\alpha}$ now has the same form as the currents in (3.487) which gives the relationships

$$L^{(11)} = M^{(11)} \tag{3.494}$$

$$L^{(21)} = L^{(12)} = M^{(12)} - \mu M^{(11)} \tag{3.495}$$

$$L^{(22)} = M^{(22)} - 2\mu M^{(12)} + \mu^2 M^{(11)} \tag{3.496}$$

A third possible definition of heat or energy current is the choice

$$\mathbf{J}_W = \mathbf{J}_E + eV\mathbf{J} = \mathbf{J}_Q + \bar{\mu}\mathbf{J} \tag{3.497}$$

Here one includes the fact that the potential $V(\mathbf{r})$ may not be constant in the sample, so that a charged particle which moves will acquire a change in potential energy. Determining the forces which go with this choice of energy current is a homework assignment.

The thermal conductivity is usually measured under conditions of no particle current $\mathbf{J} = 0$. From (3.481), this condition leads to

$$\frac{e}{T}\nabla V + \nabla\left(\frac{\mu}{T}\right) = \frac{M^{(12)}}{M^{(11)}}\nabla\left(\frac{1}{T}\right) \tag{3.498}$$

which for the energy current gives

$$\mathbf{J}_E = \nabla\left(\frac{1}{T}\right)\left[M^{(22)} - \frac{(M^{(12)})^2}{M^{(11)}}\right] \tag{3.499}$$

The M tensors are being treated as scalar quantities, which is permissible if all the forces are in the same symmetry direction. Since the thermal conductivity K is usually written as

$$\mathbf{J}_E = -K\nabla T \tag{3.500}$$

$$K = \frac{1}{T^2}\left[M^{(22)} - \frac{(M^{(12)})^2}{M^{(11)}}\right] \tag{3.501}$$

Equations (3.475) and (3.476) and the constraint $\partial S/\partial t > 0$ can be used to show that the right-hand side is always positive (Kubo, 1959). If there is no particle flow, then $\mathbf{J}_Q = \mathbf{J}_E$. The

thermal conductivity can be found from (3.487) in a similar fashion. The restriction that no current flows gives:

$$\frac{1}{T}\nabla\bar{\mu} = \frac{L^{(12)}}{L^{(11)}}\nabla\left(\frac{1}{T}\right) \tag{3.502}$$

The heat current and thermal conductivity are

$$\mathbf{J}_Q = \nabla\left(\frac{1}{T}\right)\left[L^{(22)} - \frac{(L^{(12)})^2}{L^{(11)}}\right] \tag{3.503}$$

$$K = \frac{1}{T^2}\left[L^{(22)} - \frac{(L^{(12)})^2}{L^{(11)}}\right] \tag{3.504}$$

It is easy to use the identities (3.495) to show that this definition is consistent with (3.501).

The electrical conductivity is usually defined when there is no temperature gradient $\nabla T = 0$ and no concentration gradient $\nabla\mu = 0$:

$$\sigma = \frac{e^2}{T}L^{(11)} \tag{3.505}$$

The coefficient $L^{(12)}$ may also be measured, since it is the thermoelectric coefficient. If a solid has a temperature gradient ΔT, no particle currents $\mathbf{J} = 0$, and no concentration gradients ($\nabla\mu = 0$), then a voltage difference ΔV is measured which is proportional to ΔT and is given by

$$\Delta V = -\frac{1}{eT}\frac{L^{(12)}}{L^{(11)}}\Delta T \tag{3.506}$$

This equation is the definition of the thermoelectric coefficient S, which is also called the thermopower:

$$S = -\frac{\Delta V}{\Delta T} = \frac{1}{eT}\frac{L^{(12)}}{L^{(11)}} \tag{3.507}$$

The thermopower S may have either sign. The sign usually indicates the sign of the charge on the conducting particle. The correlation function is also related to the Peltier coefficient (Barnard, 1972).

3.9.3. Correlation Functions

The coefficients $M^{(ij)}$, and $L^{(ij)}$, are correlation functions of current operators. They are expressed by formulas similar to those of the preceding section. One wishes to obtain the measurable currents $\mathbf{J}_i$ by evaluating the expectation value of the operator $\mathbf{j}_i$ of the same quantity:

$$\mathbf{J}_i = \langle \mathbf{j}_i \rangle \tag{3.508}$$

The derivation is identical to the steps used to get (3.428). One introduces a term F into the Hamiltonian, and this term causes the change in the density matrix. In direct analogy with (3.428), the result is

$$\mathbf{J}_i = \int_0^\infty dt e^{-st}\int_0^\beta d\beta' \mathrm{Tr}\left[\rho_0 \frac{\partial}{\partial t}F(-t - i\beta')\mathbf{j}_i(\mathbf{r})\right] \tag{3.509}$$

The only remaining step is to make the proper choice for $\partial F/\partial t$. It is the time rate of change of the energy of the system. At least it is the dissipative part brought on by the transport and forces. It is the heat production or equivalently the irreversible rate of change of entropy S:

$$\frac{\partial F}{\partial t} = \frac{dQ}{dt} = T\frac{\partial S}{\partial t} = \frac{1}{\beta}\sum_l \mathbf{j}_l \cdot \mathbf{X}_l \tag{3.510}$$

The small $\mathbf{j}_i$ symbols are for current operators, not their averages. Using this result for $\partial F/\partial t$ in (3.509),

$$\mathbf{J}_i = -\frac{1}{\beta}\int_0^\infty dte^{-st}\int_0^\beta d\beta' \mathrm{Tr}\left[\rho_0 \sum_l j_{l,\alpha}(-t - i\beta')\mathbf{X}_{l,\alpha}\mathbf{j}_i(\mathbf{r})\right] \tag{3.511}$$

Since the forces $\mathbf{X}_l$ are constants, they may be removed from the integrals. Then $\mathbf{J}_i$ is proportional to $\mathbf{X}_l$, and the constants of proportionality are just the transport coefficient:

$$Z_{\beta\alpha}^{(il)} = -\frac{1}{\beta}\int_0^\infty dte^{-st}\int_0^\beta d\beta' \mathrm{Tr}\,[\rho_0 j_{l,\alpha}(-t - i\beta')j_{i,\beta}(\mathbf{r})] \tag{3.512}$$

Equation (3.512) is a very simple result. The transport coefficient between force $\mathbf{X}_l$ and current $\mathbf{J}_i$ is just the current–current correlation function of $\mathbf{j}_l$ and $\mathbf{j}_i$. For example, the coefficients in (3.487) are

$$L_{\alpha\beta}^{(11)} = -\frac{1}{\beta}\int_0^\infty dte^{-st}\int_0^\beta d\beta' \mathrm{Tr}\,[\rho_0 j_\alpha(-t - i\beta')j_\beta] \tag{3.513}$$

$$L_{\alpha\beta}^{(12)} = -\frac{1}{\beta}\int_0^\infty dte^{-st}\int_0^\beta d\beta' \mathrm{Tr}\,[\rho_0 j_{Q,\alpha}(-t - i\beta')j_\beta] \tag{3.514}$$

$$L_{\alpha\beta}^{(22)} = -\frac{1}{\beta}\int_0^\infty dte^{-st}\int_0^\beta d\beta' \mathrm{Tr}\,[\rho_0 j_{Q,\alpha}(-t - i\beta')j_{Q,\beta}] \tag{3.515}$$

The $M^{(ij)}$ coefficients in (3.481) are similar; the only difference is that $\mathbf{j}_E$ is used instead of $\mathbf{j}_Q$. Since $\mathbf{j}_Q = \mathbf{j}_E - \mu\mathbf{j}$, it is easy to show that the relations (3.495) are automatically satisfied. It is also easy to prove the Onsager relations. First one writes the correlation function in the $|n\rangle$ and $|m\rangle$ notation of (3.396):

$$Z^{(il)} = \pi\sum_{nm} e^{-\beta E_n}\langle n|\, j_l|m\rangle\langle m|\, j_i|n\rangle\delta(E_n - E_m) \tag{3.516}$$

$$Z^{(li)} = \pi\sum_{nm} e^{-\beta E_n}\langle n|\, j_i|m\rangle\langle m|\, j_l|n\rangle\delta(E_n - E_m) \tag{3.517}$$

These two expressions can be shown to be identical by interchanging dummy summation variables n and m and using $E_n = E_m$.

These correlation functions can also be evaluated by using the Matsubara formalism. One just evaluates the correlation function

$$Z^{(il)} = \frac{1}{\omega_n\beta}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau j_l(\tau)j_i(0)\rangle \tag{3.518}$$

by the usual S-matrix techniques. After a suitable result is obtained, one gets the retarded function by letting $i\omega_n \to \omega + i\delta$ and then zero frequency by letting $\omega = 0$. It was shown in Sec. 3.7 that these steps are identical to the evaluation of (3.512).

PROBLEMS

1. Prove the following moments of the spectral functions:

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \omega^2 n_F(\omega) A(\mathbf{p}, \omega) = -\langle [H, C_{\mathbf{p}\sigma}^\dagger][H, C_{\mathbf{p}\sigma}] \rangle \tag{3.519}$$

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \omega n_B(\omega) B(\mathbf{q}, \omega) = \langle A(\mathbf{q})[H, A(-\mathbf{q})] \rangle \tag{3.520}$$

Evaluate these moments for noninteracting particles and phonons.

2. Take the Fourier transform of the correlation function (3.1). Define its spectral function as $C(\mathbf{p}, \omega)$. Show that

$$A(\mathbf{p}, \omega) = (1 + e^{-\beta\omega}) C(\mathbf{p}, \omega) \tag{3.521}$$

What is the result when (3.1) is time-ordered?

3. Let $U(t)$ be the density operator $\rho(\mathbf{q}, t)$ of wave vector $\mathbf{q}$. Prove that

$$S(\mathbf{q}) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} n_B(\omega) S(\mathbf{q}, \omega) \tag{3.522}$$

where $S(\mathbf{q}) = \langle \rho(\mathbf{q})\rho(-\mathbf{q}) \rangle$ is the pair distribution function and $S(\mathbf{q}, \omega)$ is the spectral function for the density–density correlation function.

4. At zero temperature,

   (a) Show that $n_{\mathbf{p}} = \int_{-\infty}^{0} (d\omega/2\pi) A(\mathbf{p}, \omega)$.
   (b) Evaluate this integral for a Lorentzian: $A(\mathbf{p}, \omega) = 2\Gamma_{\mathbf{p}}/[(\omega - \xi_{\mathbf{p}})^2 + \Gamma_{\mathbf{p}}^2]$.
   (c) Plot $n_{\mathbf{p}}$ vs. $\xi_{\mathbf{p}}$ for cases (1) $\Gamma_{\mathbf{p}} = \Gamma_0 =$ constant, and (2) $\Gamma_{\mathbf{p}} = \Gamma_0 |\xi_{\mathbf{p}}| / \sqrt{\Gamma_0^2 + \xi_{\mathbf{p}}^2}$.

5. Let $C_{\mathbf{p}}^\dagger$ be a creation operator describing a spinless boson whose number is conserved; this model is sometimes applied to ^{4}He atoms if one can ignore the internal structure (excitations, degrees of freedom). The particles will have a chemical potential μ. Define

$$\mathcal{G}(\mathbf{p}, \tau) = -\langle T_\tau e^{\beta(\Omega - K)} C_{\mathbf{p}}(\tau) C_{\mathbf{p}}^\dagger(0) \rangle \tag{3.523}$$

   (a) Show that $\mathcal{G}^{(0)}(\mathbf{p}, i\omega) = 1/(i\omega - \xi_{\mathbf{p}})$.
   (b) What is the definition of the retarded function?

6. Prove the following general theorem for particles which are fermions or bosons:

$$G_{\text{ret}}(\mathbf{p}, t) = -i\Theta(t) \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A(\mathbf{p}, \omega) e^{-i\omega t} \tag{3.524}$$

7. Evaluate the thermodynamic average for the case where the operators represent bosons

$$\left\langle \sum_{\mathbf{k}} a_{\mathbf{k}}^\dagger(\tau) a_{\mathbf{k}}(\tau) \sum_{\mathbf{p}} a_{\mathbf{p}}^\dagger(\tau') a_{\mathbf{p}}(\tau') \right\rangle \tag{3.525}$$

by doing it (a) directly and (b) using Wick's theorem. Both should give the same answer. Terms with $\mathbf{p} = \mathbf{k}$ have to be treated carefully.

8. Use the rules for constructing diagrams to write down the two-phonon self-energy diagrams of the electron: $\Sigma^{(2)}(\mathbf{p}, ip_n)$.

9. The two diagrams in Fig. 3.12 contribute to the electron self-energy and arise from the electron–electron interaction. Use the rules for constructing diagrams, and write down these contributions in the Matsubara formalism. However, do not evaluate them, i.e., do neither the frequency summations nor the wave vector integrals.

10. How is (3.215) altered if the internal line is a Coulomb interaction?

11. Derive the summation formulas (3.217) and (3.218). Also evaluate

$$\frac{1}{\beta}\sum_{i\omega_n}\mathcal{D}^{(0)}(\mathbf{q}, i\omega_n)\mathcal{D}^{(0)}(\mathbf{k}, i\omega_n + iq_m) \tag{3.526}$$

12. Evaluate the Matsubara frequency summation for

$$\frac{1}{\beta}\sum_{iq_m}\mathcal{G}^{(0)}(\mathbf{p}, ip_n + iq_m)\mathcal{D}(\mathbf{q}, iq_m) \tag{3.527}$$

13. Prove that

$$\Omega_0 = \lim_{\tau\to 0}\left\{\frac{2}{\beta}\sum_{\mathbf{p}, ip_n} e^{ip_n\tau}\ln[\mathcal{G}^{(0)}(\mathbf{p}, ip_n)]\right\} \tag{3.528}$$

where $\mathcal{G}^{(0)}$ is for electrons and Ω_0 is the free-electron part of the thermodynamic potential (Luttinger and Ward, 1960).

14. Consider the terms W_8 for the electron–phonon interaction. How many times do each of the terms appear: U_4^2, U_2^4, U_6U_2, $U_4U_2^2$?

15. Evaluate $W_1(\mathbf{p}, \tau)$ in (3.314) using $\hat{V}(\tau_1)$ as the electron–electron interaction. Which terms would not make sense if one tried to write this as $\exp(F_1)$?

16. Show that at nonzero temperatures the real part of the conductivity can also be written as

$$\mathrm{Re}[\sigma_{\alpha\beta}(\omega)] = \frac{1 - e^{-\beta\omega}}{2\omega v}\int_{-\infty}^{\infty} dt e^{it\omega}\langle j_\alpha^\dagger(\mathbf{q}, t) j_\beta(\mathbf{q}, 0)\rangle \tag{3.529}$$

17. Find the forces $\mathbf{X}_i$ which go with the currents $\mathbf{J}$, $\mathbf{J}_W$ in (3.497). What are the Kubo formulas for the transport coefficients?

18. Show that the relations (3.495) are consistent with (3.512).

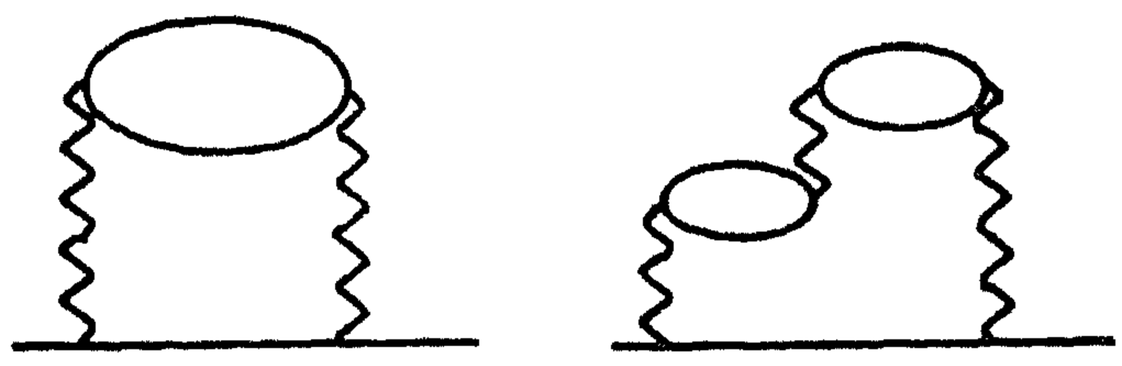


FIGURE 3.12

19. Show that the second-order phonon self-energy, from the electron–phonon interaction, may be written as $\Pi(\mathbf{q}, \omega) = M_{\mathbf{q}}^2 P^{(1)}(\mathbf{q}, \omega)$:

$$p^{(1)}(\mathbf{q}, \omega) = 2\int \frac{d^3p}{(2\pi)^3} n_{\mathbf{p}} \left(\frac{1}{\omega - \varepsilon_{\mathbf{p}} + \varepsilon_{\mathbf{p}+\mathbf{q}} + i\delta} - \frac{1}{\omega + \varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + i\delta} \right)$$

Evaluate this expression by doing the wave vector integrals. Assume a free-electron gas of Fermi momentum k_F at zero temperature. Give both the real and imaginary parts of $P^{(1)}$.

20. Derive the change in the thermodynamic potential when the electron self-energy is $\Sigma(\eta) = \eta\Sigma_0$.

21. Derive the change in the thermodynamic potential when the electron self-energy is $\Sigma(\eta) = \eta^2\Sigma_0$.

22. The thermodynamic potential Ω, internal energy U, and entropy S are related by $\Omega = U - TS - \mu N$, where

$$S = -\left(\frac{d\Omega}{dT}\right)_{\mu, V} \tag{3.530}$$

$$C_V = T\left(\frac{dS}{dT}\right)_{\mu, V} \tag{3.531}$$

Derive the expressions for S, U, and C_V for the free-electron gas. Show that the specific heat is proportional to T at low T, and find the constant of proportionality.

23. Derive the relationship between the total internal and external electric fields by solving Maxwell's equations.

24. Show that Wick's theorem is invalid when H_0 contains particle–particle interactions. As an example, evaluate the correlation function W below for a system of two states. Do it exactly and also by Wick's theorem.

$$H_0 = \sum_{\alpha=1}^{2} \xi_\alpha C_\alpha^\dagger C_\alpha + U n_1 n_2 \tag{3.532}$$

$$W = \sum_{\alpha\beta} \langle T_\tau C_\alpha^\dagger(\tau_1) C_\alpha(\tau_1) C_\beta^\dagger(\tau_2) C_\beta(\tau_2) \rangle \tag{3.533}$$

25. Derive the equilibrium real-time Green's functions in (3.327) starting from (3.322).

26. In one dimension find the Wigner distribution function for the simple harmonic oscillator of mass m and frequency ω_0. In center-of-mass coordinates show that its spectral function is given in terms of Laguerre polynomials

$$A(k, \omega; X) = 4\pi e^{-\Delta} \sum_n (-1)^n L_n(2\Delta)\delta[\omega - \omega_0(n + \tfrac{1}{2})]$$

$$\Delta = \frac{X^2}{l^2} + k^2 l^2, \quad l^2 = \frac{1}{m\omega_0}$$

27. Prove the following identity:

$$\Sigma_{\text{ret}}(\mathbf{k}, \omega; T) = -\int_{-\infty}^{\infty} \frac{d\omega'}{\omega - \omega' + i\delta} [\Sigma^>(\mathbf{k}, \omega'; T) - \Sigma^<(\mathbf{k}, \omega'; T)]$$

This result is easy to prove in equilibrium. Can you also prove it for systems out of equilibrium?

Chapter 4

Exactly Solvable Models

Every many-body theorist should be knowledgeable about the available exactly solvable models. First, there are not many of them. Second, they are useful for gaining insight into many-particle systems. If the problem to be solved can be related to an exactly solvable one, however vaguely, one can usually gain some insight.

Several models are presented here to introduce the discussion of Green's functions. Usually, the model is first solved without Green's functions and again with them. Models which are exactly solvable may be solved by a variety of techniques.

The models presented here are only a small sampling of the possible exactly solvable ones. Others have been presented by Schweber (1961). Many one-dimensional models may be solved exactly (Lieb and Mattis, 1966; Baxter, 1982; Mattis, 1993).

4.1. POTENTIAL SCATTERING

A very simple problem is an impurity potential $V(\tau)$ in an otherwise free-particle system. All other interactions are ignored except that of the free particles with the impurity, which is assumed to be at the origin. The potential is assumed to have no internal structure: spin, excited states, etc. It is a simple function of position r of the particle from the origin and is spherically symmetric.

The wave functions ψ_λ and energy ε_λ for each particle may be obtained by solving a one-electron Schrödinger equation:

$$H\psi_\lambda = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\psi_\lambda = \varepsilon_\lambda\psi_\lambda \tag{4.1}$$

The solution may be accomplished by numerical means if the Schrödinger equation is not solvable analytically.

In many-body theory, the impurity problem is usually encountered as a scattering center. The free-particle states are plane waves,

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{\nu}}\sum_{\mathbf{k}} C_{\mathbf{k}\sigma}e^{i\mathbf{k}\cdot\mathbf{r}} \tag{4.2}$$

The spin index σ is dropped, since spin effects are unimportant. The Hamiltonian is expressed as operators

$$H = \sum_{\mathbf{k}} \varepsilon_k C_{\mathbf{k}}^\dagger C_{\mathbf{k}} + \frac{1}{\nu} \sum_{\mathbf{k}\mathbf{k}'} V_{\mathbf{k}\mathbf{k}'} C_{\mathbf{k}}^\dagger C_{\mathbf{k}'} \tag{4.3}$$

$$V_{\mathbf{k}\mathbf{k}'} = \int d^3 r V(r) e^{-i\mathbf{r}\cdot(\mathbf{k}-\mathbf{k}')} = V(\mathbf{k}-\mathbf{k}') \tag{4.4}$$

The last term is the potential scattering of the free particles. The object is to diagonalize the Hamiltonian (4.3). Of course, the solutions are given by (4.1). However, the problem is not entirely solved. Since the equation is second order, it has two solutions. These may be chosen as the ingoing and outgoing waves. Alternately, these two may be combined to give standing waves. These choices must be related to the scattering problem implied in (4.3).

The discussion will benefit from the knowledge that the solution is of the form (4.1). Consider the boundary conditions on the wave function. The integral equation for the wave function is:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r}) + \sum_{\mathbf{k}'} \frac{\phi_{\mathbf{k}'}(\mathbf{r})}{\varepsilon_k - \varepsilon_{k'}} \int d^3 r' \phi_{\mathbf{k}'}^*(\mathbf{r}') V(r') \psi_{\mathbf{k}}(\mathbf{r}') \tag{4.5}$$

$$\phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\nu}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{4.6}$$

$$\varepsilon_k = \frac{\hbar^2 k^2}{2m} \tag{4.7}$$

This form of the integral equation is valid for the free-particle states with energy ε_k. The Schrödinger equation (4.1) may also have bound states. They shall not be discussed explicitly. They obey an equation similar to (4.5), but the energy is changed to the binding energy $\varepsilon_k \to -\varepsilon_B$ ($\varepsilon_B > 0$), and the $\phi_{\mathbf{k}}$ term on the right is absent.

The first step is to prove that (4.5) is equivalent to (4.1). Operate on both sides of the equation by $H_0 - \varepsilon_k$,

$$(H_0 - \varepsilon_k)\psi_{\mathbf{k}} = (H_0 - \varepsilon_k)\phi_{\mathbf{k}} + \sum_{\mathbf{k}'} \frac{(H_0 - \varepsilon_k)\phi_{\mathbf{k}'}(\mathbf{r})}{\varepsilon_k - \varepsilon_{k'}} \times \int d^3 r' \phi_{\mathbf{k}'}^*(\mathbf{r}') V(r') \psi_{\mathbf{k}}(\mathbf{r}')$$

and use the fact that

$$(H_0 - \varepsilon_k)\phi_{\mathbf{k}} = -\left(\frac{\hbar^2}{2m}\nabla^2 + \varepsilon_k\right)\phi_{\mathbf{k}} = 0 \tag{4.8}$$

to get

$$(H_0 - \varepsilon_k)\psi_{\mathbf{k}} = -\sum_{\mathbf{k}'} \phi_{\mathbf{k}'}(\mathbf{r}) \int d^3 r' \phi_{\mathbf{k}'}^*(\mathbf{r}') V(r') \psi_{\mathbf{k}}(\mathbf{r}') \tag{4.9}$$

The completeness relation for the summation over the set of states is

$$\sum_{\mathbf{k}'} \phi_{\mathbf{k}'}(\mathbf{r}) \phi_{\mathbf{k}'}^*(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \tag{4.10}$$

$$(H_0 - \varepsilon_k)\psi_{\mathbf{k}} = -V(r)\psi_{\mathbf{k}}(\mathbf{r}) \tag{4.11}$$

If the potential is transferred to the left, the final equation is

$$(H_0 + V - \varepsilon_k)\psi_{\mathbf{k}}(\mathbf{r}) = 0 \tag{4.12}$$

which is the desired answer. The above equation is precisely (4.1). It has been shown that (4.5) is equivalent to the usual Schrödinger equation. Next discuss boundary conditions. The differential equation (4.1) is second order, so there are two independent solutions. As a mathematical problem, these two solutions may be combined in any possible way, depending on the choice of boundary conditions. As a physics problem the two solutions are combined to give a desirable physical situation. One choice is to have the wave function a standing wave. This choice leads to the reaction matrix equation. Other choices are to have the wave function an incoming wave or an outgoing wave, which leads to T-matrix theory. From scattering theory and causality a natural choice is the outgoing wave. It has the asymptotic form $\exp(ikr)/r$ at large distance. This choice of phase [rather than $\exp(-ikr)/r$] arises from the physics convention that the time development is $\exp(-i\omega t)$. The integral equation (4.5) is a convenient starting point for this discussion of boundary conditions, since the various choices of standing, outgoing, or ingoing waves are determined only by the complex part of the energy denominator. The factor is $\varepsilon_k - \varepsilon_{k'}$, for standing waves, so that the principal part is chosen for the denominator. The factor is $\varepsilon_k - \varepsilon_{k'} + i\delta$ for outgoing waves, and $\varepsilon_k - \varepsilon_{k'} - i\delta$ for ingoing waves.

4.1.1. Reaction Matrix

Here the energy denominator is chosen to be real and given by the principal part. The free-particle Green's function is then defined as

$$G_0(\mathbf{k}, \mathbf{r} - \mathbf{r}') = P\sum_{\mathbf{k}'} \frac{\phi_{\mathbf{k}'}(\mathbf{r})\phi^*_{\mathbf{k}'}(\mathbf{r}')}{\varepsilon_k - \varepsilon_{k'}} = \frac{P}{\nu}\sum_{\mathbf{k}'} \frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{\varepsilon_k - \varepsilon_{k'}} \tag{4.13}$$

$$= P\int \frac{d^3k'}{(2\pi)^3} \frac{e^{i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{\varepsilon_k - \varepsilon'_k} \tag{4.14}$$

The last step on the right takes $\nu \to \infty$ and changes the summation into an integration over wave vectors. The integral is standard and gives

$$G_0(\mathbf{k}, \mathbf{r} - \mathbf{r}') = \pi\rho(k)\frac{\cos[k|\mathbf{r} - \mathbf{r}'|]}{k|\mathbf{r} - \mathbf{r}'|} \tag{4.15}$$

A factor $\rho(k)$ has been introduced which is the density of states of the particles:

$$\rho(k) = \int \frac{d^3k'}{(2\pi)^3}\delta(\varepsilon_k - \varepsilon_{k'}) = \frac{mk}{2\hbar^2\pi^2} \tag{4.16}$$

For particles with spin, this factor is the density of states per spin configuration. In electron systems, for example, the net density of states is twice this result when spin degeneracy is considered. The Green's function may be expanded as a function of $\mathbf{r}$ and $\mathbf{r}'$:

$$G_0(\mathbf{k}, \mathbf{r} - \mathbf{r}') = \pi\rho(k)\sum_l (2l+1)P_l(\hat{r}\cdot\hat{r}')j_l(kr_<)\eta_l(kr_>) \tag{4.17}$$

where the notation $r_<$ and $r_>$ means the lesser and greater of r and r', respectively. The $j_l(kr)$ and $\eta_l(kr)$ are spherical Bessel functions of the first and second kind, and the $P_l(\cos\theta)$ are Legendre functions. These results may be put into the integral equation (4.5):

$$\psi_{\mathbf{k}}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r}) + \int d^3r' G_0(\mathbf{k}, \mathbf{r} - \mathbf{r}')V(r')\psi_{\mathbf{k}}(\mathbf{r}') \tag{4.18}$$

At this point it is convenient to reduce the integral equation to angular momentum components l. A plane wave may be expanded as

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_l (2l+1)i^l P_l(\hat{k}\cdot\hat{r})j_l(kr) \tag{4.19}$$

A similar expansion is used for the actual wave function:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_l (2l+1)i^l P_l(\hat{k}\cdot\hat{r})R_l(kr) \tag{4.20}$$

The radial function $R_l(kr)$ is the quantity to be determined. It satisfies a radial Schrödinger equation of the form

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial R}{\partial r} - \frac{l(l+1)}{r^2}R\right] + [V(r) - \varepsilon_k]R = 0 \tag{4.21}$$

which does not determine the boundary conditions. They are obtained by substituting these forms into (4.18). Each angular momentum component is selected by multiplying the equation by $P_l(\hat{k}\cdot\hat{r})$ and then integrating over all spatial angles. One uses the fact that

$$\int d\Omega_r P_l(\hat{k}\cdot\hat{r})P_m(\hat{r}\cdot\hat{p}) = \frac{4\pi}{2l+1}\delta_{lm}P_l(\hat{k}\cdot\hat{p}) \tag{4.22}$$

to reduce the equation down to one which involves only the same angular momentum component:

$$\begin{aligned} R_l(kr) &= j_l(kr) + 4\pi^2\rho(k)\int_0^\infty r'^2 dr' j_l(kr_<)\eta_l(kr_>)V(r')R_l(kr') \\ &= j_l(kr) + 4\pi^2\rho(k)\left[\eta_l(kr)\int_0^r r'^2 dr' j_l(kr')V(r')R_l(kr')\right. \\ &\quad \left. + j_l(kr)\int_r^\infty r'^2 dr' \eta_l(kr')V(r')R_l(kr')\right] \end{aligned} \tag{4.23}$$

It is important that the potential is spherically symmetric. Otherwise the scattering term would mix angular momentum components, which usually makes the equation much harder to solve.

The solution is examined in the limit as $kr \to \infty$. From the differential equation (4.21) it can be shown that the radial wave function must asymptotically approach the value

$$\lim_{kr\to\infty} R_l(kr) \to \frac{C_l(k)}{kr}\sin\left[kr + \delta_l(k) - \frac{l\pi}{2}\right] \tag{4.24}$$

The prefactor C_l is determined below. The asymptotic limit of the integral equation (4.23) is

$$\begin{aligned} \lim R_l(kr) &\to j_l(jr) + D_l(k)\eta_l(kr) \\ &\to \frac{\sin(kr - l\pi/2)}{kr} - D_l\frac{\cos(kr - l\pi/2)}{kr} \end{aligned} \tag{4.25}$$

where

$$D_l(k) = 4\pi^2 \rho(k) \int_0^\infty r^2 dr j_l(kr) V(r) R_l(kr) \tag{4.26}$$

The potential $V(r)$ is assumed to be of short range: It falls off faster than r^{-2} at large distances. This assumption is necessary for the integral in (4.26) to be well defined. The scattering is then described by a phase shift $\delta_l(k)$ which depends on angular momentum and wave vector. The two asymptotic expansions (4.24) and (4.25) must be identical, which is accomplished by setting

$$D_l = -\tan(\delta_l) = 4\pi^2 \rho(k) \int_0^\infty r^2 dr j_l(kr) V(r) R_l(kr). \tag{4.27}$$

The complicated integral just defines the tangent of the phase shift. Now (4.25) becomes

$$\begin{aligned} \lim_{kr\to\infty} R_l(kr) &\to \frac{1}{kr\cos(\delta)}[\cos(\delta)\ \sin(kr - l\pi/2) + \sin(\delta)\ \cos(kr - l\pi/2)] \\ &= \frac{\sin(kr + \delta_l - l\pi/2)}{kr\cos(\delta_l)} \end{aligned} \tag{4.28}$$

The normalization coefficient in (4.24) is

$$C_l = \frac{1}{\cos(\delta_l)} \tag{4.29}$$

When solving the radial wave function (4.21), the solution is obtained which is well behaved at the origin. This solution is followed outward in r until the region is reached where $V(r) \approx 0$ and the centrifugal barrier $\hbar^2 l(l+1)/(2mr^2)$ is small. Then the solution has the form (4.24) with $C_l = 1/\cos(\delta_l)$. The wave function is now properly normalized at large r, and by following it back toward the origin it is normalized everywhere. These steps provide the proper solution to the scattering equation (4.5) which was the starting point in the calculation.

The reaction matrix is defined as the quantity

$$R_{\mathbf{k}'\mathbf{k}} = \int d^3 r \phi^*_{\mathbf{k}'}(\mathbf{r}) V(r) \psi_{\mathbf{k}}(\mathbf{r}) \tag{4.30}$$

It may be expanded in angular momentum states by using the expansions for the wave function and the plane wave. The angular integrals give

$$R_{\mathbf{k}'\mathbf{k}} = 4\pi \sum_l (2l+1) P_l(\hat{k}\cdot\hat{k}') R_l(k', k) \tag{4.31}$$

where the radial integral is

$$R_l(k', k) = \int_0^\infty r^2 dr j_l(k'r) V(r) R_l(kr) \tag{4.32}$$

It is defined for the general case where $k' \neq k$. Of course, if they happen to be equal, then the answer is just (4.27):

$$R_l(k, k) = -\frac{\tan(\delta_l)}{4\pi^2 \rho(k)} = -\frac{\hbar^2\ \tan(\delta_l)}{2mk} \tag{4.33}$$

where the density of states is $\rho(k)$. The reaction matrix is only proportional to $\tan(\delta)$ for the diagonal terms where $k' = k$. Otherwise, one must do the integral in (4.32).

The reaction matrix obeys an integral equation which is deduced by putting (4.30) into the integral equation (4.5) for $\psi_{\mathbf{k}}(\mathbf{r})$

$$R_{\mathbf{k}'\mathbf{k}} = V_{\mathbf{k}'\mathbf{k}} + P\sum_{\mathbf{k}_1} \frac{V_{\mathbf{k}'\mathbf{k}_1} R_{\mathbf{k}_1\mathbf{k}}}{\varepsilon_k - \varepsilon_{k_1}} \tag{4.34}$$

This integral equation is frequently found in scattering problems. Now it is solved easily. For each angular momentum state, the one-particle radial equation (4.21) is solved with the boundary condition that the asymptotic limit goes to (4.24). This wave function is used in (4.32), which determines the answer. It is a simple procedure to solve (4.34). The important point is that the wave functions must be normalized correctly. The designation of principal parts on the energy denominator in (4.34) specifies the unique way this normalization is to be done.

4.1.2. T Matrix

The other common choice of boundary conditions uses outgoing waves, which is accomplished by adding an infinitesimal complex part $i\delta$ to the energy denominator in (4.5). The integral equation is

$$\psi_{\mathbf{k}}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r}) + \int d^3r' G_0(\mathbf{k}, \mathbf{r} - \mathbf{r}')V(r')\psi_{\mathbf{k}}(\mathbf{r}') \tag{4.35}$$

where the Green's function for outgoing waves is

$$G_0(\mathbf{k}, \mathbf{r} - \mathbf{r}') = \int \frac{d^3k'}{(2\pi)^3} \frac{e^{-i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')}}{\varepsilon_k - \varepsilon_{k'} + i\delta} \tag{4.36}$$

This integral is evaluated to give

$$G_0(\mathbf{k}, R) = -\pi\rho(k)\frac{e^{ikR}}{kR} = -\frac{\pi\rho(k)}{kR}[\cos(kR) + i\sin(kR)] \tag{4.37}$$

The cosine term was evaluated earlier in (4.17). The sine term is

$$\frac{\sin(k|\mathbf{r} - \mathbf{r}'|)}{k|\mathbf{r} - \mathbf{r}'|} = \sum_l (2l + 1)P_l(\hat{r}\cdot\hat{r}')j_l(kr)j_l(kr') \tag{4.38}$$

These results can be put into the integral equation (4.35) to give an equation for the wave function. Averaging over angles gives for each angular momentum state:

$$\tilde{R}_l(kr) = j_l(kr) + 4\pi^2\rho(k)\int_0^\infty r'^2 dr'[j_l(kr_<)\eta_l(kr_>) - ij_l(kr)j_l(kr')]V(r')\tilde{R}_l(kr') \tag{4.39}$$

The radial part of the wave function is denoted by $\tilde{R}_l$. It will be a different wave function than found for the standing-wave boundary conditions. The difference between the two is only a

different choice of normalization coefficient $C_l(k)$ in (4.24). This different choice is what is to be determined. The above integral equation may be rewritten as

$$\begin{aligned}\tilde{R}_l(kr) = j_l(kr)\Big[1 &+ 4\pi^2\rho(k)\int_r^\infty r'^2 dr' \eta_l(kr')V(r')\tilde{R}_l(kr') \\ &-4\pi^2 i\rho(k)\int_0^\infty r'^2 dr' j_l(kr')V(r')\tilde{R}_l(kr')\Big] \\ &+ 4\pi^2\rho(k)\eta_l(kr)\int_0^r r'^2 dr' j_l(kr')V(r')\tilde{R}_l(kr')\end{aligned} \tag{4.40}$$

Then one can take the limit $kr \to \infty$ and obtain

$$\lim_{kr\to\infty} \tilde{R}_l(kr) \to j_l(kr)[1 - i\tilde{D}_l] + \tilde{D}_l(k)\eta_l(kr) \tag{4.41}$$

$$\tilde{D}_l(k) = 4\pi^2\rho(k)\int_0^\infty r^2 dr j_l(kr)V(r)\tilde{R}_l(kr) \tag{4.42}$$

This equation has to have the form (4.24) because the radial wave function $\tilde{R}_l(kr)$ also obeys the differential equation (4.21). The result (4.41) has that form if

$$\tilde{D}_l(k) = -e^{i\delta_l}\ \sin(\delta_l) \tag{4.43}$$

so that the factor multiplying $j_l(kr)$ is

$$\begin{aligned}1 - i\tilde{D}_l(k) &= 1 + ie^{i\delta_l}\ \sin(\delta_l) = 1 + \frac{e^{i\delta_l}}{2}(e^{i\delta_l} - e^{-i\delta_l}) \\ &= e^{i\delta_l}\ \cos(\delta_l)\end{aligned} \tag{4.44}$$

For the asymptotic limit

$$\begin{aligned}\lim_{kr\to\infty} \tilde{R}_l(kr) &\to e^{i\delta_l}[\,j_l(kr)\ \cos(\delta_l) - \eta_l(kr)\ \sin(\delta_l)] \\ &= \frac{e^{i\delta_l}}{kr}[\sin(kr - l\pi/2)\ \cos(\delta_l) + \cos(kr - l\pi/2)\ \sin(\delta_l)] \\ &= \frac{e^{i\delta_l}}{kr}\ \sin(kr + \delta_l - l\pi/2)\end{aligned} \tag{4.45}$$

which is indeed the right form, (4.24). For outgoing wave boundary conditions, the proper choice of normalization coefficient is

$$C_l(k) = e^{i\delta_l(k)} \tag{4.46}$$

One should proceed by solving (4.21) and insist that it have this asymptotic limit. The easiest way is to first solve it for a real wave function and make this real wave function go to $[\sin(kr + \delta - l\pi/2)]/kr$ [which determines $\delta_l(k)$]. Then one can multiply the wave function everywhere by the phase factor $\exp[i\delta_l(k)]$.

The T matrix is defined as

$$T_{\mathbf{k}'\mathbf{k}} = \int d^3r \phi^*_{\mathbf{k}'}(\mathbf{r}) V(r) \psi_{\mathbf{k}}(\mathbf{r})$$
$$= 4\pi \sum_l (2l+1) P_l(\hat{k}\cdot\hat{k}') T_l(k', k) \tag{4.47}$$

$$T_l(k', k) = \int_0^\infty r^2 dr j_l(k'r) V(r) \tilde{R}_l(kr) \tag{4.48}$$

The original integral equation for the wave function (4.35) may be used to generate the T-matrix equation

$$T_{\mathbf{k}'\mathbf{k}} = V_{\mathbf{k}'\mathbf{k}} + P \sum_{\mathbf{k}_1} \frac{V_{\mathbf{k}'\mathbf{k}_1} T_{\mathbf{k}_1\mathbf{k}}}{\varepsilon_k - \varepsilon_{k_1} + i\delta} \tag{4.49}$$

This equation is often encountered in scattering problems. The solution is easy. One solves Schrödinger's equation (4.1) for the wave function $\psi_{\mathbf{k}}(\mathbf{r})$. Only the radial part of the wave function is difficult, since one must solve (4.21). The solution is normalized by insisting that the radial wave function have the form (4.24) at long distance from the potential. The coefficient is $C_l = \exp[i\delta_l(k)]$. These radial solutions are used in (4.47), which then gives an exact solution for the T-matrix equation. This result differs from the reaction matrix result only in the choice of the coefficient C_l which multiplies the radial wave function. The angular momentum components of the two scattering functions are related by the ratio of these normalization coefficients:

$$\tilde{R}_l(kr) = e^{i\delta_l} \cos(\delta_l) R_l(kr) \tag{4.50}$$

$$T_l(k', k) = e^{i\delta_l} \cos(\delta_l) R_l(k', k) \tag{4.51}$$

The T matrix is a complex quantity since it was defined with a complex $\exp[i\delta_l]$ phase factor. The diagonal T matrix is

$$T_l(k, k) = \frac{-1}{4\pi^2 \rho(k)} e^{i\delta_l} \sin[\delta_l(k)] = -\frac{1}{2mk} e^{i\delta_l} \sin[\delta_l] \tag{4.52}$$

We repeat again that this simple form is valid only when $k = k'$. The imaginary part of the diagonal T matrix is

$$-\mathrm{Im}[T_l(k, k)] = \frac{\sin^2(\delta_l)}{4\pi^2 \rho(k)} \tag{4.53}$$

$$-\mathrm{Im}[T_{\mathbf{k}\mathbf{k}}] = \frac{1}{\pi\rho(k)} \sum_l (2l+1) \sin^2[\delta_l(k)] \tag{4.54}$$

which is related to the total scattering cross section for the potential:

$$\sigma = \frac{4\pi}{k^2} \sum_l (2l+1) \sin^2[\delta_l(k)] \tag{4.55}$$

One can also prove the *optical theorem*

$$-2\,\mathrm{Im}[T_{\mathbf{k}\mathbf{k}}] = 2\pi \int \frac{d^3k}{(2\pi)^3} |T_{\mathbf{k}'\mathbf{k}}|^2 \delta(\varepsilon_k - \varepsilon_{k'}) \tag{4.56}$$

It is important to realize that the reaction matrix is *not* identical to the real part of the T matrix. Either of these two different quantities may be selected as the energy shift of a particle interacting with a potential.

4.1.3. Friedel's Theorem

There are several important theorems involving phase shifts. These theorems are conveniently proved by letting the solid have a spherical shape of radius R, with the impurity located at the center of the sphere. The distance R is very large, and eventually the limit is taken of $R \to \infty$. A finite size sample is used since we are going to count nodes in the wave function. Start counting at the surface of the sphere and move inward. It is assumed, for convenience, that the wave functions vanish at the surface of the sphere.

In the absence of the potential, the solutions which are regular at the origin have a radial part $j_l(kr)$. The condition that they vanish at the surface of the sphere is

$$j_l(k_n R) = 0 \tag{4.57}$$

Since small values of l are important, say $l < 5$ for most applications, and since R is very large, an adequate approximation at large argument is

$$j_l(k_n R) \to \frac{1}{k_n R} \sin\left(k_n R - \frac{l\pi}{2}\right) \tag{4.58}$$

The condition for the vanishing of the wave function at the surface is

$$k_n R = \left(n + \frac{l}{2}\right)\pi \tag{4.59}$$

For each value of l, there is a solution for each additional integer n, where each additional integer has a solution with an extra node in the wave function. Similarly, the continuum wave functions (i.e., for $\varepsilon_k > 0$) in the presence of the impurity at large r is

$$R_l(kr) \to \frac{C_l}{kr} \sin\left(kr + \delta_l - \frac{l\pi}{2}\right) \tag{4.60}$$

The solution is obtained for each k_n which satisfies the equation

$$k_n R + \delta_l(k_n) = \left(n + \frac{l}{2}\right)\pi \tag{4.61}$$

Count the additional particle states in the presence of the impurity. The number of states dn between k and $k + dk$ is

$$\frac{dn}{dk} = \frac{R}{\pi} + \frac{1}{\pi}\frac{d\delta_l}{dk} \tag{4.62}$$

The first term R/π is just what one would have without the impurity. The extra states δn from the impurity are given by the formula

$$\frac{d\delta n}{dk} = \frac{1}{\pi}\frac{d\delta_l}{dk} \tag{4.63}$$

The quantity on the right $(d/dk)(\delta_l(k)/\pi)$ is interpreted as the change in the number of particle states caused by the impurity. For example, if the potential is repulsive so that particles are pushed away from the impurity region, then $d\delta_l/dk$ will be negative. Similarly,

an attractive potential will draw particles inward, so that $d\delta_l/dk$ is positive. Many potentials have $d\delta_l/dk$ positive for some wave vectors and negative for others. The quantity $d\delta_l/dk$ is the change in the density of states for each angular momentum state, each magnetic quantum number m_l, and each spin quantum number m_s. The total change in the density of states is obtained by summing over all these quantum numbers:

$$\frac{d}{dk}N = \frac{d}{dk}\sum_{m_s m_l l}\left(\frac{\delta_{lm_s m_l}}{\pi}\right) \tag{4.64}$$

In a metal, the electron states are occupied up to the Fermi wave vector k_F. The Friedel sum rule (Friedel, 1952, 1953) is obtained by integrating up to the Fermi wave vector,

$$Z = \int_0^{k_F} dk\left(\frac{dN}{dk}\right) = \sum_{m_s m_l l}\left(\frac{\delta_{lm_s m_l}(k_F)}{\pi}\right) \tag{4.65}$$

where Z is the charge on the impurity. It is assumed that the phase shift is zero at $k = 0$. Levinson's theorem is that the phase shift is zero at $k = 0$ unless there are bound states, in which case it is π times the number of bound states. The Friedel sum rule is a statement of charge neutrality. In a static electron gas, there are no long-range Coulomb potentials of the form r^{-1}. Instead, the electron gas is displaced in the vicinity of an impurity charge. The displaced electronic charge exactly cancels the impurity charge. For example, an impurity of positive valence Z has electrons drawn in, until an extra Z of electrons, or eZ of charge, surrounds the impurity. This extra charge is called a screening charge. It is shown in Fig. 4.1(a). The phase shifts are calculated for the screened potential of impurity plus screening charge. If the potential is repulsive, electrons are depleted around the impurity, so that charge neutrality is maintained. This depletion is shown in Fig. 4.1(b). The Friedel sum rule is a statement of charge neutrality: the change in electron charge around an impurity is exactly equal in magnitude and opposite in sign to the charge of the impurity. A neutral impurity would have a Friedel sum of zero. The sum rule does not mean that all the phase shifts vanish. Generally, some are plus, some are minus, and all vary with the wave vector. Only the summation at the Fermi energy is zero. Another way to express that the screening charge equals the impurity charge is

$$Z = 4\pi\int_0^{\infty} r^2 dr[\rho(r) - \rho_0] \tag{4.66}$$

where ρ_0 is the equilibrium charge density in the metal.

The charge densities in Fig. 4.1 are schematic. The method of actually calculating these curves is discussed in Chapter 5. All such curves, when calculated correctly, have the oscillations in the charge density at large distance. These variations are called *Friedel oscillations*. Their magnitude has been exaggerated in Fig. 4.1, since they are small in amplitude. But they occur in real solids and have been observed by several techniques. The

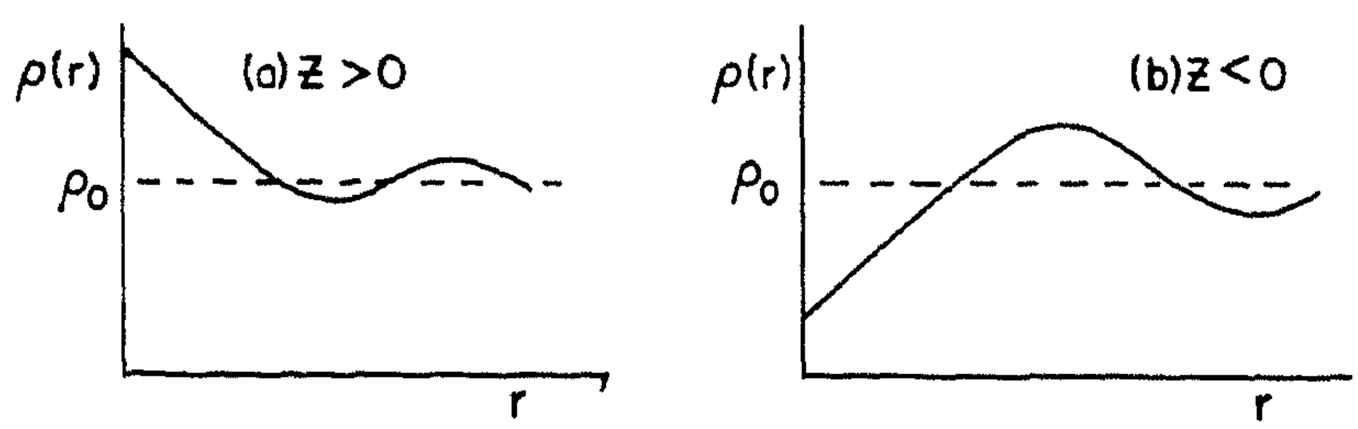


FIGURE 4.1 Electron density $\rho(r)$ when an impurity charge of Z is put at the origin. (a) For $Z > 0$, (b) for $Z < 0$.

Friedel oscillations may be derived by taking the asymptotic limit of the change in charge density,

$$\rho(r) - \rho_0 = 2\int_{k<k_F} \frac{d^3k}{(2\pi)^3}[|\psi_{\mathbf{k}}(\mathbf{r})|^2 - |\phi_{\mathbf{k}}(\mathbf{r})|^2] \tag{4.67}$$

$$\lim_{r\to\infty}[\rho(r) - \rho_0] = \frac{8\pi}{(2\pi)^3 r^2}\sum_l (2l+1)\int_0^{k_F} dk[\sin^2(kr + \delta_l - l\pi/2) - \sin^2(kr - l\pi/2)] \tag{4.68}$$

and

$$\begin{aligned}&[\sin^2(kr + \delta_l - l\pi/2) - \sin^2(kr - l\pi/2)]\\ &\quad = \tfrac{1}{2}[\cos(2kr - l\pi) - \cos(2kr + 2\delta_l - l\pi)]\end{aligned} \tag{4.69}$$

The wave vector integral is difficult because the phase shifts $\delta_l(k)$ depend on k. An approximate answer is derived by writing this dependence as $\delta_l(k) = \delta_l(k_F) + (k - k_F)(d\delta/dk)$. The k integral is then elementary and the r-dependent part is

$$\begin{aligned}&(-1)^l \int_0^{k_F} dk\left\{\cos(2kr) - \cos\left[2k\left(r + \frac{d\delta}{dk}\right) + 2\delta_l - 2k_F\frac{d\delta}{dk}\right]\right\}\\ &\quad = -\frac{(-1)^l}{2r}\{\sin(2k_F r) - \sin[2k_F r + 2\delta_l(k_F)]\}\end{aligned} \tag{4.70}$$

and the change in density is

$$\lim_{r\to\infty}[\rho(r) - \rho_0] = \frac{1}{4\pi^2 r^3}\sum_l (2l+1)\ \sin[\delta_l(k_F)] \times \cos[2k_F r + \delta_l(k_F)] + O\left(\frac{1}{r^4}\right) \tag{4.71}$$

At large distances, the changes in charge density oscillate with a period of $2k_F$ and decrease in amplitude as r^{-3}. This asymptotic equation is independent of the nature of the impurity. The impurity determines only the values for the phase shift $\delta_l(k_F)$.

The Friedel sum rule is believed to be exact in real systems. Langer and Ambegaokar (1961) have shown it to be valid even in an interacting many-particle system. If one knows the exact impurity potential and the exact screening charge profile, the Friedel sum rule will be valid. In practice, a calculation of an impurity potential is done by pseudopotential or other means, and it is screened by a good dielectric function. The phase shifts are evaluated numerically and it is found that the Friedel sum rule errs by a small percentage. This error is no fault of the rule but the choice of potential or screening function. In fact, the theoretical system is not charge neutral. Usually the potential is adjusted slightly, for example, by altering a screening length, to force the Friedel sum rule to be satisfied. Then one has a consistent model of the neutral system, and the phase shifts are probably reasonably accurate.

Quite often the phase shifts do not depend on m_l or m_s, but they strongly depend on l. Then one can represent the answer for electrons as

$$Z = \frac{2}{\pi}\sum_l (2l+1)\delta_l(k_F) \tag{4.72}$$

which is the way that it is usually presented. The factor of 2 is spin degeneracy and $(2l+1)$ is orbital degeneracy. The quantity $(2/\pi)(2l+1)\delta_l(k_F)$ is interpreted as the amount of screening charge in the angular momentum channel l.

In real solids, the impurity may occupy the site ordinarily occupied by a host atom. Then the normal host atom would have its own set of phase shifts $\delta_l^{(h)}$ which characterize the

normal metal. For example, if Mg^{2+} is a substitutional impurity in sodium metal, it could be at a site where a Na^+ is usually placed. The Na^+ ion has its own phase shifts $\delta_l^{(h)}$, and the Mg^{2+} ion has its own phase shifts $\delta_l^{(i)}$. If one repeats the arguments leading to (4.72), the Friedel sum rule is given by the difference of the two sets of phase shifts:

$$Z = \frac{2}{\pi}\sum_l (2l+1)[\delta_l^{(i)}(k_F) - \delta_l^{(h)}(k_F)] \tag{4.73}$$

Note that the host phase shifts do not obey the Friedel sum rule.

Another exact result may be obtained in terms of phase shifts. It is the total energy of the impurity as caused by its interactions with the surrounding electrons in a metal. The theorem is due to Fumi (1955) and relates the total energy E_i to an energy integral over the phase shifts:

$$E_i = -\int_0^{E_F} dE \sum_{lm_lm_s} \frac{\delta_l(E)}{\pi} \tag{4.74}$$

$$= -\frac{\hbar^2}{m}\int_0^{k_F} kdk \sum_{lm_lm_s} \frac{\delta_l(k)}{\pi} \tag{4.75}$$

The energy integral starts at the bottom of the conduction band. It is assumed there are no bound states. The presence of bound states will change the answer. One can also use the second form of (4.75), which includes just the sum over wave vector states for the conduction electrons.

Fumi's theorem is proved in the following way. The Hamiltonian is the kinetic energy plus the potential energy. Well outside the impurity, or its screening charge, the potential is zero, and the Hamiltonian only has the kinetic energy term. Here the energy must just be the summation of the kinetic energies of all the particles, which is a discrete summation over the states allowed in the sphere of radius R:

$$E = \sum_{lmn} \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2}{2mR^2}\sum_{lmn}\left[\left(n + \tfrac{1}{2}l\right)\pi - \delta_l\right]^2 \tag{4.76}$$

The desired quantity is actually the change in kinetic energy, so the result without the impurity is subtracted:

$$E_i = \frac{\hbar^2}{2mR^2}\sum_{lm_lm_sn}\{[(n+\tfrac{1}{2}l)\pi - \delta_l]^2 - [(n+\tfrac{1}{2}l)\pi]^2\} \tag{4.77}$$

$$\approx -\frac{\hbar 2}{mR}\sum_{lm_lnm_s} k_n\delta_l(k_n) \tag{4.78}$$

Next let the radius of the sphere go to infinity, which changes the wave vector summation to a continuous integration. This change and the result are

$$\frac{1}{R}\sum_n k_n\delta_l(k_n) \to \frac{1}{\pi}\int_0^{k_F} kdk\delta_l(k) \tag{4.79}$$

$$E_i = -\frac{\hbar^2}{m\pi}\sum_{lm_lm_s}\int_0^{k_F} kdk\delta_l(k) \tag{4.80}$$

which proves the theorem for the conduction band states. The result can be extended to include bound states. They rarely occur in three dimensions, but are common in lower dimensions.

4.1.4. Impurity Scattering

How are Green's functions used to describe impurity scattering? They do not work very well for describing the scattering of a particle from a single impurity. In Dyson's equation, the change in energy of the particle is calculated from the interactions. One impurity in a large system of volume ν changes the energy of a particle only by order $1/\nu$. This result assumes, of course, that the particle does not get bound by the impurity. A free-particle state has a density $|\phi_{\mathbf{k}}|^2 = 1/\nu$, so that a local impurity changes the energy only by $1/\nu$. To alter the energy, one must add N_i impurities, where $N_i/\nu = n_i$ as $\nu \to \infty$. Then the particle energy gets changed as a function of the concentration n_i of impurities. The isolated impurity case should be studied by the following procedure: one solves the self-energy as a function of n_i and then takes the limit $n_i \to 0$. Terms of order $O(n_i)$ are from an electron interaction with single impurities.

The potential which scatters the electron is taken as a summation of impurity potentials:

$$\bar{V}(\mathbf{r}) = \sum_j V(\mathbf{r} - \mathbf{R}_j) \tag{4.81}$$

The Fourier transform is $\bar{V}(\mathbf{q})$:

$$\bar{V}(\mathbf{q}) = \rho_i(\mathbf{q}) V(\mathbf{q}) \rho(-\mathbf{q}) \tag{4.82}$$

where $\rho(-\mathbf{q})$ and $\rho_i(\mathbf{q})$ are the particle densities of electrons and impurities, respectively:

$$\rho(\mathbf{q}) = \sum_{k\sigma} C^\dagger_{\mathbf{k}+\mathbf{q},\sigma} C_{\mathbf{k}\sigma} \tag{4.83}$$

$$\rho_i(\mathbf{q}) = \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{4.84}$$

The expansion of the S matrix encounters products of the impurity density operator:

$$f_n(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n) = \langle \rho_i(\mathbf{q}_1)\rho_i(\mathbf{q}_2)\cdots\rho_i(\mathbf{q}_n)\rangle \tag{4.85}$$

These products must be averaged. This average is not taken over temperature. Instead, it is averaged over the possible positions which the impurities may have in the solid. Usually it is assumed that the impurities are randomly located and that there is no correlation between their positions. This method of impurity averaging was suggested by Kohn and Luttinger (1957). Their result will be derived by first examining the results for a small number of operators. The first term is $n = 1$:

$$f_1(\mathbf{q}) = \left\langle \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \right\rangle = N_i \delta_{\mathbf{q}=0} \tag{4.86}$$

If the $\mathbf{R}_j$ are located randomly, this sum is zero unless $\mathbf{q} = 0$, and then the sum gives the number of impurities N_i. The second case is $n = 2$:

$$\left\langle \sum_{ij} e^{i\mathbf{q}_1\cdot\mathbf{R}_i + i\mathbf{q}_2\cdot\mathbf{R}_j} \right\rangle = \left\langle \sum_{i=j} e^{i(\mathbf{q}_1+\mathbf{q}_2)\cdot\mathbf{R}_j} + \sum_{i\neq j} e^{i\mathbf{q}_1\cdot\mathbf{R}_i + i\mathbf{q}_2\cdot\mathbf{R}_j} \right\rangle \tag{4.87}$$

The first term is zero unless $\mathbf{q}_1 + \mathbf{q}_2 = 0$, since it is the same average as in f_1. The second term is zero unless both $\mathbf{q}_1$ and $\mathbf{q}_2$ equal zero:

$$f_2(\mathbf{q}_1, \mathbf{q}_2) = N_i \delta_{\mathbf{q}_1+\mathbf{q}_2=0} + N_i(N_i - 1)\delta_{\mathbf{q}_1=0}\delta_{\mathbf{q}_2=0} \tag{4.88}$$

The general result found by Kohn and Luttinger is

$$\begin{aligned} f_n(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n) = {} & N_i \delta_{\Sigma_i \mathbf{q}_i=0} + N_i(N_i - 1) \sum_m \delta_{\Sigma_{j=1}^m \mathbf{q}_j=0} \delta_{\Sigma_{j=m+1}^n \mathbf{q}_j=0} \\ & + N_i(N_i - 1)(N_i - 2)\delta_{\Sigma_{\mathbf{q}=0}} \delta_{\Sigma_{\mathbf{q}=0}} \delta_{\Sigma_{\mathbf{q}=0}} + \cdots \end{aligned} \tag{4.89}$$

In terms with products of several delta functions $\delta_{\Sigma_{\mathbf{q}=0}}$, one must take all possible combinations of distributing the $\mathbf{q}_j$ among the delta functions. The final result is obtained by generalizing the low-order examples above. The result for $n = 3$ is

$$\begin{aligned} f_3(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = {} & N_i \delta_{\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3} + N_i(N_i - 1)(\delta_{\mathbf{q}_1}\delta_{\mathbf{q}_2+\mathbf{q}_3} + \delta_{\mathbf{q}_2}\delta_{\mathbf{q}_1+\mathbf{q}_3} + \delta_{\mathbf{q}_3}\delta_{\mathbf{q}_1+\mathbf{q}_2}) \\ & + N_i(N_i - 1)(N_i - 2)\delta_{\mathbf{q}_1}\delta_{\mathbf{q}_2}\delta_{\mathbf{q}_3} \end{aligned} \tag{4.90}$$

The next step is to learn to draw Feynman diagrams with these impurity averages. Since $N_i > 10^{10}$ for real systems, this expression is approximated by

$$f_n(\mathbf{q}_1, \ldots, \mathbf{q}_n) = N_i \delta_{\Sigma_\mathbf{q}} + N_i^2 \delta_{\Sigma_\mathbf{q}} \delta_{\Sigma_\mathbf{q}} + N_i^3 \delta_{\Sigma_\mathbf{q}} \delta_{\Sigma_\mathbf{q}} \delta_{\Sigma_\mathbf{q}} + \cdots \tag{4.91}$$

There occurs the combinations of the function

$$N_i \delta_{\Sigma_\mathbf{q}} \tag{4.92}$$

This factor is interpreted as the particle scattering from a single impurity. The factor of $\delta_{\Sigma_\mathbf{q}}$ states that momentum is conserved for the particle while scattering from the impurity. In Fig. 4.2(a), the impurity is represented by an $\times$, and the solid line is the particle. The dashed lines represent impurity interactions $V(\mathbf{q}_j)$. Momentum conservation requires the momenta from a single impurity to sum to zero. Diagrams with two factors of (4.92) involve the scattering from two impurities, as shown in Figs 4.2(b) and 4.2(c).

In Fig. 4.2(b), the scattering from the two impurities interfere—the momentum lines cross each other. Such interference occurs when the impurities are nearby, and such diagrams are important when the concentration of impurities is large. In Fig. 4.2(c), the two scattering events are disconnected, since the interaction lines do not overlap. Each connected diagram gives a separate self-energy contribution. Edwards (1958) has shown that one has the Dyson equation

$$\mathcal{G}(\mathbf{p}, ip_n) = \frac{1}{ip_n - \xi_p - \Sigma(\mathbf{p}, ip_n)} \tag{4.93}$$

where the self-energy $\Sigma(\mathbf{p}, ip_n)$ from impurity scattering contains all the connected contributions. The diagram in Fig. 4.2(b) is connected, so it contributes a term to Σ. The self-energy

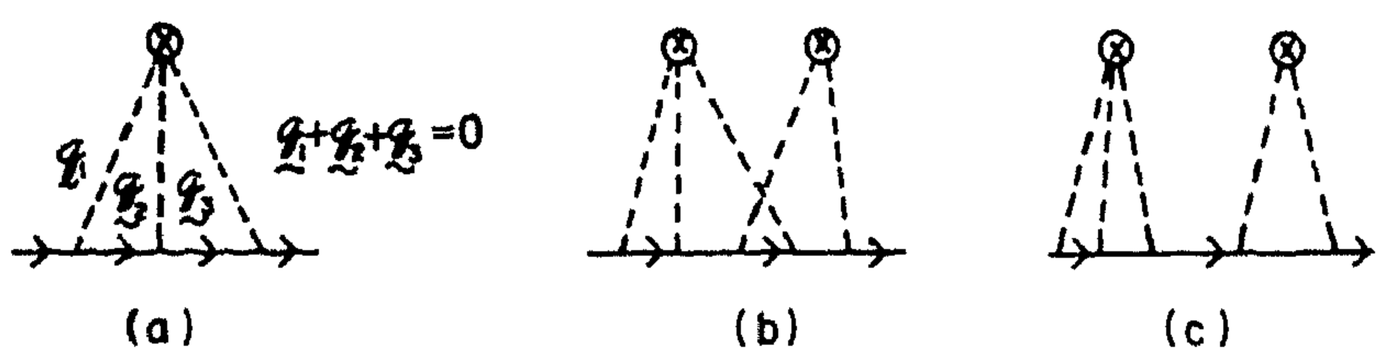


FIGURE 4.2

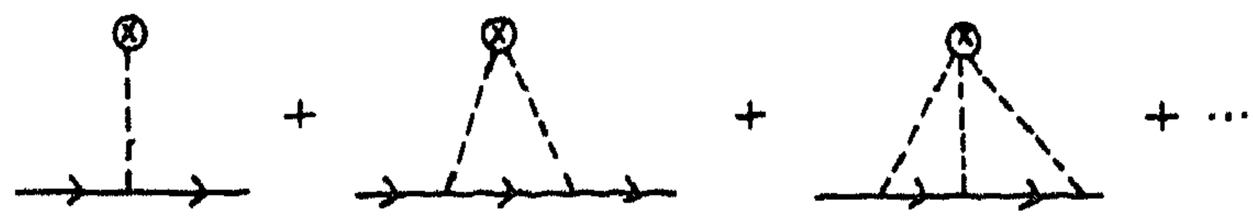


FIGURE 4.3

diagrams for scattering from a single impurity are shown in Fig. 4.3. They are the terms which are important when the concentration of impurities is small. The first self-energy diagram in Fig. 4.3 gives a contribution

$$\Sigma^{(1)}(\mathbf{p}, ip_n) = \frac{N_i}{\nu} \sum_{\mathbf{q}} \delta_{\mathbf{q}=0} V(\mathbf{q}) = n_i V(0) \tag{4.94}$$

The second diagram in the series has a particle line as an intermediate state. This self-energy term is

$$\Sigma^{(2)}(\mathbf{p}, ip_n) = \frac{N_i}{\nu^2} \sum_{\mathbf{q}_1 \mathbf{q}_2} \delta_{\mathbf{q}_1+\mathbf{q}_2} V(\mathbf{q}_1) V(\mathbf{q}_2) \mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}_1, ip_n) \tag{4.95}$$

which may also be written as ($\nu \to \infty$)

$$\Sigma^{(2)}(\mathbf{p}, ip_n) = n_i \int \frac{d^3p'}{(2\pi)^3} \frac{V(\mathbf{p}-\mathbf{p}')V(\mathbf{p}'-\mathbf{p})}{ip_n - \xi_{\mathbf{p}'}} \tag{4.96}$$

There is no summation over internal frequency variables. The impurity is considered rigid and cannot absorb or transfer energy to the particle which is scattering. Consequently, all the internal lines have the same energy ip_n as the initial particle. This result is not assumed but instead is derived directly from the Green's function expansion. For example, the above self-energy term arose from second-order terms and had the τ integrals

$$\begin{aligned} &\int_0^\beta d\tau e^{ip_n\tau} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \mathscr{G}^{(0)}(\mathbf{p}, \tau-\tau_1)\mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}_1, \tau_1-\tau_2)\mathscr{G}^{(0)}(\mathbf{p}, \tau_2) \\ &\quad = \mathscr{G}^{(0)}(\mathbf{p}, ip_n)^2 \mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}_1, ip_n) \end{aligned} \tag{4.97}$$

The fact that all electron lines have the same energy comes directly from the τ integrals. It is true for an arbitrary order of diagram—all particle Green's functions have the same energy ip_n. As stated before, there is no other excitation in the model to which energy may be transferred.

The third term in the series in Fig. 4.3 is

$$\Sigma^{(3)}(\mathbf{p}, ip_n) = n_i \int \frac{d^3p_1}{(2\pi)^3} \frac{d^3p_2}{(2\pi)^3} \frac{V(\mathbf{p}-\mathbf{p}_1)V(\mathbf{p}_1-\mathbf{p}_2)V(\mathbf{p}_2-\mathbf{p})}{(ip_n - \xi_{\mathbf{p}_1})(ip_n - \xi_{\mathbf{p}_2})}$$

At this point, one may deduce the nth term by inspection. It is

$$\begin{aligned} \Sigma^{(n)}(\mathbf{p}, ip_n) = n_i \int &\frac{d^3p_1}{(2\pi)^3} \cdots \frac{d^3p_n}{(2\pi)^3} V(\mathbf{p}-\mathbf{p}_1)V(\mathbf{p}_1-\mathbf{p}_2)\cdots V(\mathbf{p}_n-\mathbf{p}) \\ &\times \Pi_{j=1}^n \mathscr{G}^{(0)}(\mathbf{p}_j, ip_n) \end{aligned} \tag{4.98}$$

Now that the nth term is evident, it is also possible to sum the series of terms and get the total self-energy which is proportional to the concentration n_i. The first step in this summation is to

Max - Planck
für Mikrostruk
Biblioth

write an integral equation for the self-energy. This integral equation is in terms of a vertex function $\Gamma(\mathbf{p}', \mathbf{p})$ which has the equation

$$\Gamma(\mathbf{p}', \mathbf{p}) = V(\mathbf{p}' - \mathbf{p}) + \int \frac{d^3p_1}{(2\pi)^3} \frac{V(\mathbf{p}' - \mathbf{p}_1)\Gamma(\mathbf{p}_1, \mathbf{p})}{ip_n - \xi_{p1}} \tag{4.99}$$

$$\begin{aligned}\Sigma(\mathbf{p}, ip_n) &= \sum_{n=1}^{\infty} \Sigma^{(n)}(\mathbf{p}, ip_n) = n_i V(0) + \int \frac{d^3p_1}{(2\pi)^3} \frac{V(\mathbf{p} - \mathbf{p}_1)\Gamma(\mathbf{p}_1, \mathbf{p})}{ip_n - \xi_{p1}} \\ &= n_i \Gamma(\mathbf{p}, \mathbf{p})\end{aligned} \tag{4.100}$$

Repeated iteration of Eq. (4.99) will generate the series of terms which give the successive diagrams in Fig. 4.3, of which (4.98) is the nth term. The series is summed by solving the integral equation. To this end, define the function

$$\pi(\mathbf{r}, \mathbf{p}) = \int \frac{d^3p_1}{(2\pi)^3} \frac{e^{i\mathbf{p}_1 \cdot \mathbf{r}}\Gamma(\mathbf{p}_1, \mathbf{p})}{ip_n - \xi_{p1}} \tag{4.101}$$

This quantity and $\Gamma(\mathbf{p}', \mathbf{p})$ are both functions of ip_n, but this dependence is not explicitly added to the notation. In terms of this new function, the self-energy (4.100) is

$$\Sigma(\mathbf{p}, ip_n) = n_i V(0) + n_i \int d^3r e^{-i\mathbf{p} \cdot \mathbf{r}} V(\tau)\pi(\mathbf{r}, \mathbf{p}) \tag{4.102}$$

To solve for the function $\pi(\mathbf{r}, \mathbf{p})$, consider the effects of the differential operator acting on it:

$$\begin{aligned}\left[-\frac{\hbar^2}{2m}\nabla^2 - \mu - ip_n\right]\pi(\mathbf{r}, \mathbf{p}) &= \int \frac{d^3p'}{(2\pi)^3} \frac{\xi_{\mathbf{p}'} - ip_n}{ip_n - \xi_{\mathbf{p}'}} e^{i\mathbf{p}' \cdot \mathbf{r}}\Gamma(\mathbf{p}', \mathbf{p}) \\ \left[-\frac{\hbar^2}{2m}\nabla^2 - \mu - ip_n\right]\pi(\mathbf{r}, \mathbf{p}) &= -\int \frac{d^3p'}{(2\pi)^3} e^{i\mathbf{p}' \cdot \mathbf{r}}\Gamma(\mathbf{p}', \mathbf{p})\end{aligned} \tag{4.103}$$

The right-hand side of this equation may be simplified. First, note that the vertex function may be written as

$$\Gamma(\mathbf{p}', \mathbf{p}) = \int d^3r e^{-i\mathbf{p}' \cdot \mathbf{r}} V(r)[e^{i\mathbf{p} \cdot \mathbf{r}} + \pi(\mathbf{r}, \mathbf{p})] \tag{4.104}$$

which follows directly from its definition (4.99). From this relation, it is easy to show that

$$\int \frac{d^3p'}{(2\pi)^3} \Gamma(\mathbf{p}', \mathbf{p}) e^{i\mathbf{p}' \cdot \mathbf{r}} = V(r)[e^{i\mathbf{p} \cdot \mathbf{r}} + \pi(\mathbf{r}, \mathbf{p})] \tag{4.105}$$

It is deduced that the function obeys an inhomogeneous differential equation:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 - \mu - ip_n + V(r)\right]\pi(\mathbf{r}, \mathbf{p}) = -V(r)e^{i\mathbf{p} \cdot \mathbf{r}} \tag{4.106}$$

This equation is solved in the following fashion. First find the solution to the following Schrödinger equation:

$$H\psi_\lambda = \left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\psi_\lambda(\mathbf{r}) = E_\lambda \psi_\lambda \tag{4.107}$$

This differential equation was discussed extensively at the beginning of this section. Take the outgoing wave boundary conditions. The reason for this choice will be clear below. In terms

of these wave functions, the special function $\pi(\mathbf{r}, \mathbf{p})$ has the following solution to the inhomogeneous equation (4.106):

$$\pi(\mathbf{r}, \mathbf{p}) = \int d^3r' e^{i\mathbf{p}\cdot\mathbf{r}'} V(r') \sum_\lambda \frac{\psi_\lambda^*(\mathbf{r}')\psi_\lambda(\mathbf{r})}{ip_n + \mu - E_\lambda} = \sum_\lambda \frac{\psi_\lambda(\mathbf{r}) T_{\lambda\mathbf{p}}^*}{ip_n + \mu - E_\lambda} \tag{4.108}$$

where the T matrix is defined in (4.47). This result for $\pi(\mathbf{r}, \mathbf{p})$ can be directly verified by operating on both sides by $[H - \mu - ip_n]$ and then using the completeness property:

$$\sum_\lambda \psi_\lambda^*(\mathbf{r}')\psi_\lambda(\mathbf{r}) = \delta^3(\mathbf{r} - \mathbf{r}') \tag{4.109}$$

The summation λ is taken over all the eigenstates of H, which include the bound states as well as the continuum states. When this result is used in (4.102), the final result is obtained for the self-energy in terms of the T matrices which were defined in (4.47):

$$\Sigma(\mathbf{p}, ip_n) = n_i V(0) + n_i \sum_\lambda \frac{|T_{\lambda\mathbf{p}}|^2}{ip_n + \mu - E_\lambda} \tag{4.110}$$

Equation (4.110) is the exact term in the self-energy which is proportional to the concentration n_i. It has been obtained by summing the set of diagrams shown in Fig. 4.3, where the particle has multiple scattering events from the same impurity. There are other terms in the self-energy which are proportional to higher powers of the concentration $(n_i)^m$. For example, one term which contributes to order n_i^2 is shown in Fig. 4.2(b). However, in the limit of dilute impurities, as $n_i \to 0$, the result (4.110) is the most important.

Our interest is usually in the retarded self-energy, which is called $\Sigma(\mathbf{p}, \omega)$, omitting the "ret" subscript:

$$\Sigma(\mathbf{p}, \omega) = n_i V(0) + n_i \sum_\lambda \frac{|T_{\lambda\mathbf{p}}|^2}{\omega - \xi_\lambda + i\delta} \tag{4.111}$$

There is one particular case which is very important: when the energy ω is set equal to ξ_p. Then the self-energy (4.111) is just n_i times the diagonal T matrix:

$$\Sigma(\mathbf{p}, \xi_p) = n_i T_{\mathbf{pp}} = -\frac{2\pi n_i}{mp} \sum_l (2l+1) e^{i\delta_l(p)} \sin[\delta_l(p)] \tag{4.112}$$

This identification is made by noting that the T-matrix equation (4.49) is identical to the vertex equation (4.99) when $ip_n \to \xi_p + i\delta$. This result is frequently used in representing the self-energy of the particle. The imaginary part of the self-energy has a simple formula:

$$-2\ \mathrm{Im}[\Sigma(\mathbf{p}, \omega)] = 2\pi n_i \sum_\lambda |T_{\lambda\mathbf{p}}|^2 \delta(\omega - \xi_\lambda) \tag{4.113}$$

The right-hand side may be further reduced in the special case $\omega = \xi_p$. If the two prior results (4.55) and (4.56) are combined, the result may be shown to be proportional to the scattering cross section:

$$-2\ \mathrm{Im}[\Sigma(\mathbf{p}, \xi_p)] = -2n_i\ \mathrm{Im}\{T_{\mathbf{pp}}\} = n_i v_p \sigma(p) \tag{4.114}$$

The imaginary part of the self-energy for $\omega = \xi_p$ is the cross section times $v_p n_i$. The factor $v_p n_i$ is just the rate at which the particle encounters the impurities. The lifetime and mean-free-path of the particle may be identified as

$$\frac{1}{\tau(p)} = n_i v_p \sigma(p) \tag{4.115}$$

$$l_p = \frac{1}{n_i \sigma(p)} = v_p \tau(p) \tag{4.116}$$

These results provide an example of the earlier assertion that the imaginary part of the retarded self-energy is related to the damping of the particle.

An interesting question is the relationship between these self-energy expressions and Fumi's theorem (4.75). The two results are quite compatible. The self-energy expressions describe the effect of the impurities on single-particle states. Fumi's theorem gives the energy change of the system. Fumi's theorem can be derived by averaging the single-particle properties. This derivation is shown in the next section.

4.1.5. Ground State Energy

The ground state energy is derived starting from the self-energy expression (4.111). When averaged correctly, this expression gives Fumi's theorem (4.75) for the energy per impurity. What is proved, say at zero temperature, is that the impurities cause a change in the ground state energy of

$$\Delta\Omega = -N_i \frac{2}{\pi m} \int_0^{k_F} p dp \sum_l (2l+1)\delta_l(p) + O(N_i^2) \tag{4.117}$$

Terms proportional to $O(N_i^2)$ give energy terms arising from interactions between impurities. The term proportional to a single power N_i is the average energy per impurity from interactions with the electrons.

The result (4.117) has an interesting history. Most of our references have been to solid state physics work such as that of Friedel and Fumi. A parallel development was occurring in nuclear theory. Brueckner *et al.* (1954) proposed that the ground state energy of the system was an average over the energy of the single-particle states. The energy per particle was taken to be the reaction matrix, so they took the energy change to be

$$\Delta\Omega = 2N_i \int \frac{d^3p}{(2\pi)^3} n_F(\xi_p) R_{\mathbf{pp}} = -N_i \frac{2}{\pi m} \int_0^{k_F} p dp \sum_l (2l+1) \tan[\delta_l(p)]$$

This equation has $\tan[\delta_l]$ instead of δ_l and so is incorrect. Fukuda and Newton (1956) showed that the correct result should be δ_l rather than $\tan[\delta_l]$. In fact, they give results identical to those of Fumi. Fukuda and Newton, and also DeWitt (1956), proved some important theorems regarding the energy (4.117). The result is not dependent on the spherical box which was used in the derivation; the same result was demonstrated for a cube and other shapes. Similarly, it does not depend on the assumption that the wave function vanished at the surface of the box. The same result is obtained for other boundary conditions. Fumi's theorem appears to be a general result.

The ground state energy is calculated using the formula (3.300) derived earlier, which contains a coupling constant integration

$$\Delta\Omega = \frac{\nu}{\beta}\sum_{ip,\sigma}\int\frac{d^3p}{(2\pi)^3}\int_0^1\frac{d\eta}{\eta}\mathcal{G}(\mathbf{p}, ip, \eta)\Sigma(\mathbf{p}, ip, \eta) \tag{4.118}$$

Since only the term proportional to n_i is needed, and since the self-energy Σ is already proportional to n_i, then replace $\mathcal{G}$ by $\mathcal{G}^{(0)}$. The difference is just self-energy terms which give higher powers in $(n_i)^m$, which are going to be neglected. The coupling constant integration is important for weighting each factor V of impurity potential (and not each factor of the T matrix). From (4.98), this self-energy is

$$\begin{aligned}\Sigma(\mathbf{p}, ip, \eta) &= n_i\left[\eta V_{\mathbf{pp}} + \eta^2\int\frac{d^3p'}{(2\pi)^3}\frac{V_{\mathbf{pp}'}V_{\mathbf{p}'\mathbf{p}}}{ip-\xi_{p'}}\right.\\ &\quad\left.+\eta^3\int\frac{d^3p'}{(2\pi)^3}\int\frac{d^3p''}{(2\pi)^3}\frac{V_{\mathbf{pp}'}V_{\mathbf{p}'\mathbf{p}''}V_{\mathbf{p}''\mathbf{p}}}{(ip-\xi_{p'})(ip-\xi_{p''})}+\dots\right]\\ \Delta\Omega &= 2N_i\int\frac{d^3p}{(2\pi)^3}\frac{1}{\beta}\sum_{ip}\mathcal{G}^{(0)}(\mathbf{p}, ip)\left[V_{\mathbf{pp}}+\frac{1}{2}\int\frac{d^3p'}{(2\pi)^3}\frac{V_{\mathbf{pp}'}V_{\mathbf{p}'\mathbf{p}}}{ip-\xi_{p'}}\right.\\ &\quad\left.+\frac{1}{3}\int\frac{d^3p'}{(2\pi)^3}\int\frac{d^3p''}{(2\pi)^3}\frac{V_{\mathbf{pp}'}V_{\mathbf{p}'\mathbf{p}''}V_{\mathbf{p}''\mathbf{p}}}{(ip-\xi_{p'})(ip-\xi_{p''})}+\dots\right]\end{aligned} \tag{4.119}$$

where the coupling constant integration has been done. There are several ways to evaluate this expression.

The first derivation will give $\tan(\delta_l)$, instead of δ_l, so that it gives the wrong answer. This derivation is included, not just for fun, but to show how simple it can be to get the wrong result. While reading this derivation, try to find the incorrect step, which leads to the wrong result. The incorrect derivation proceeds by doing the summation on Matsubara frequencies in (4.119) term by term:

$$\frac{1}{\beta}\sum_{ip}\mathcal{G}^{(0)}(\mathbf{p}, ip) = n_F(\xi_{\mathbf{p}}) \tag{4.120}$$

$$\frac{1}{\beta}\sum_{ip}\mathcal{G}^{(0)}(\mathbf{p}, ip)\mathcal{G}^{(0)}(\mathbf{p}', ip) = \frac{n_F(\xi_{\mathbf{p}}) - n_F(\xi_{\mathbf{p}'})}{\xi_{\mathbf{p}} - \xi'_{\mathbf{p}}} \tag{4.121}$$

The nth term is found from the contour integral,

$$\oint\frac{dz}{2\pi i}n_F(z)\Pi_{j=1}^n\frac{1}{z-\xi_j} = 0 \tag{4.122}$$

where the contour is taken at infinity so that it includes all the poles. The contour integral leads to the identity

$$\frac{1}{\beta}\sum_{ip}\Pi_{j=1}^n\mathcal{G}^{(0)}(\mathbf{p}_j, ip) = \sum_{l=1}^n n_F(\xi_l)\Pi_{j=1,j\neq l}^n\mathcal{G}^{(0)}(\mathbf{p}_j, \xi_l) \tag{4.123}$$

The expression for the ground state energy becomes

$$\Delta\Omega = 2N_i\left\{\int\frac{d^3p}{(2\pi)^3}n_F(\xi_p)V_{\mathbf{pp}} + \frac{1}{2}\int\frac{d^3p}{(2\pi)^3}\int\frac{d^3p'}{(2\pi)^3}\right.$$
$$\left.\times V^2_{\mathbf{pp}'}[n_F(\xi_p)\mathscr{G}^{(0)}(\mathbf{p}',\xi_p) + n_F(\xi_{p'})\mathscr{G}^{(0)}(\mathbf{p},\xi_{p'})] + \cdots\right\} \tag{4.124}$$

There are two terms proportional to V^2. They are identical, after changing around the dummy variables of integration, so this cancels the $\frac{1}{2}$ factor. Similarly, there are three terms with three powers of V which are all equal, which cancels the $\frac{1}{3}$ factor. The $1/n$ factor is canceled in the nth term:

$$\Delta\Omega = 2N_i\int\frac{d^3p}{(2\pi)^3}n_F(\xi_p)\left[V_{\mathbf{pp}} + \int\frac{d^3p}{(2\pi)^3}\int\frac{d^3p'}{(2\pi)^3}\frac{V^2_{\mathbf{pp}'}}{\xi_p - \xi_{p'}}\right.$$
$$\left. + \int\frac{d^3p'}{(2\pi)^3}\int\frac{d^3p''}{(2\pi)^3}\frac{V_{\mathbf{pp}'}V_{\mathbf{p}'\mathbf{p}''}V_{\mathbf{p}''\mathbf{p}}}{(\xi_p - \xi_{p'})(\xi_p - \xi_{p''})}\cdots\right] \tag{4.125}$$

The series in brackets is just the reaction matrix equation (4.34), so that

$$\Delta\Omega = 2N_i\int_0^\beta d\tau n_F(\xi_p)R_{\mathbf{pp}} = -\frac{2N_i}{\pi m}\int_0^{k_F} pdp\sum_l(2l+1)\ \tan[\delta_l(p)] \tag{4.126}$$

The derivation gives $\tan\delta_l$, which is the wrong answer. Did you find the wrong step? The reaction matrix equation (4.126) contains a P for principal parts. It is an instruction to omit the term where the energy denominator vanishes. The T-matrix equation does not have this instruction, because one is supposed to include such terms. The derivation mishandled these terms. The "theorem" (4.123) is correct only as long as the ξ_j are different; it is incorrect if any are alike. The correct result, valid for all cases, is obtained by evaluating the contour integral (4.122) by drawing a branch cut along the real axis. The summation over Matsubara frequencies is the difference between the integration along the top and bottom of the branch cut (see Sec. 3.5). When taking the imaginary part of the right-hand side, it is possible to get products of odd numbers of delta functions. For $n = 3$ the result is

$$\frac{1}{\beta}\sum_{ip}\frac{1}{(ip-\xi_1)(ip-\xi_2)(ip-\xi_3)}$$
$$= \frac{n_F(\xi_1)}{(\xi_1-\xi_2)(\xi_1-\xi_3)} + \frac{n_F(\xi_2)}{(\xi_2-\xi_1)(\xi_2-\xi_3)} + \frac{n_F(\xi_3)}{(\xi_3-\xi_2)(\xi_3-\xi_1)}$$
$$-\pi^2 n_F(\xi_1)\delta(\xi_1-\xi_2)\delta(\xi_1-\xi_3) \tag{4.127}$$

The last term contains the product of delta functions. This term was omitted in the first derivation. Similar correction terms, with odd powers of delta functions, occur in the fourth and all higher terms in the perturbation series (4.124). An examination of these terms shows that they are generating a series of reaction matrices:

$$\Delta\Omega = 2N_i\int\frac{d^3p}{(2\pi)^3}n_F(\xi_p)\left\{R_{\mathbf{pp}} - \frac{1}{3}\int\frac{d^3p_1d^3p_2}{(2\pi)^6}R_{\mathbf{pp}_1}R_{\mathbf{p}_1\mathbf{p}_2}R_{\mathbf{p}_2\mathbf{p}}\delta(\xi_p-\xi_{p_1})\delta(\xi_p-\xi_{p_2}) + \cdots\right\} \tag{4.128}$$

The wave vector integrals may be done term by term. The d^3p_n integral eliminates the delta function $\delta(\xi_p - \xi_{p_n})$ and makes $p_n = p$. The angular integrations force all angular momentum l components to be the same. These steps derive the series

$$\Delta\Omega = \frac{2N_i}{\pi m}\int_0^{k_F} pdp \sum_l (2l+1)$$
$$\times \{[4\pi^2\rho R_l(p,p)] - \tfrac{1}{3}[4\pi^2\rho R_l(p,p)]^3 + \tfrac{1}{5}[4\pi^2\rho R_l(p,p)]^5 - \cdots\} \qquad (4.129)$$

The factor $4\pi^2\rho R_l(p,p) = -\tan[\delta_l(p)]$. The series is just that for the $\tan^{-1}(x) = x - x/3 + x/5\ldots$ where $x = -\tan(\delta)$. The ground state energy per particle as given by Fumi's theorem is:

$$\Delta\Omega = -\frac{2N_i}{\pi m}\int_0^{k_F} pdp \sum_l (2l+1)\delta_l(p) \qquad (4.130)$$

This derivation contains two messages. The first is that Fumi's theorem with $\delta_l(p)$ is correct and may be obtained by using the conventional formulas for the ground state energy of the system. The second message is that the energy of the system is not just a simple average of the single-particle energies, such as $\Sigma_{\mathbf{p}} n_F(\xi_p)R_{\mathbf{pp}}$ or $\Sigma_{\mathbf{p}} n_F(\xi_p)T_{\mathbf{pp}}$. Both of these guesses are incorrect.

A more elegant derivation of Fumi's theorem is possible by following the techniques of Langer and Ambegaokar (1961). They avoided evaluating the series (4.119) on a term-by-term basis. Instead, they used operator manipulations to attain the above result.

4.2. LOCALIZED STATE IN THE CONTINUUM

In this section the following Hamiltonian is solved exactly:

$$H = \varepsilon_c b^\dagger b + \sum_{\mathbf{k}}[\varepsilon_{\mathbf{k}} c_{\mathbf{k}}^\dagger c_{\mathbf{k}} + A_{\mathbf{k}}(c_{\mathbf{k}}^\dagger b + b^\dagger c_{\mathbf{k}})] \qquad (4.131)$$

It describes a localized state of fixed energy ε_c and operators b and $b^\dagger$. The localized state will be called the impurity, and it is assumed only one exists. There is a continuous set of states of energy $\varepsilon_{\mathbf{k}}$ with operators $c_{\mathbf{k}}$ and $c_{\mathbf{k}}^\dagger$. This set of states could have a finite bandwidth, as often occurs in tight-binding models in solids, or else it could be a free-particle model. The last term in the Hamiltonian includes the mixing between these two kinds of states. It contains processes whereby the continuum particle hops onto the impurity ($b^\dagger c_{\mathbf{k}}$) and where the particle hops off the impurity into the continuum ($c_{\mathbf{k}}^\dagger b$). If the particles have spin, it is assumed that the hopping on and off the impurity preserves the particle spin state. The spin never changes and is unimportant, so its dependence is suppressed in all subscripts and labels. If the hopping particle could change its spin orientation, the problem would become harder.

Since the Hamiltonian is quadratic in operators, its solution is equivalent to diagonalizing a matrix. The solution may be obtained in this fashion, but here it is done analytically. Since the Hamiltonian is quadratic, the statistics are irrelevant since the same eigenvalues are obtained for fermions or bosons. It is also irrelevant whether there are one or many particles in the system. For a Hamiltonian which contains only quadratic operators of fermions or bosons, one just diagonalizes the Hamiltonian to find the eigenstates, and then all eigenstates behave independently.

This model Hamiltonian was introduced simultaneously by Anderson (1961) and Fano (1961). Anderson applied it to solid-state physics, while Fano used it in atomic spectra. It tends to be called the Anderson model or the Fano model depending on whether the speaker is a solid state or atomic physicist. Here it is called the *Fano–Anderson model.* As explained in Chapter 1, it is related to the famous Anderson model, which is (4.131) plus another term.

The Hamiltonian will first be solved without using Green's functions. Afterwards, the Green's function solution will be given. The nature of the solution depends critically on whether the energy ε_c is within the band of states $\varepsilon_{\mathbf{k}}$. If the continuous band of states $\varepsilon_{\mathbf{k}}$ is confined to the range

$$w_1 \leq \varepsilon_{\mathbf{k}} \leq w_2 \tag{4.132}$$

then the solution depends on whether ε_c is also within this range. Actually, this statement is incorrect. Because of the interactions with the continuous band of states, the energy of the localized state is altered to a new value $\tilde{\varepsilon}_c$. Of course, since the Hamiltonian is not yet solved, the renormalized energy $\tilde{\varepsilon}_c$ is not yet known. Jumping ahead, the final result is that the new energy is

$$\tilde{\varepsilon}_c = \varepsilon_c + \sum_{\mathbf{k}} \frac{A_{\mathbf{k}}^2}{\tilde{\varepsilon}_c - \varepsilon_{\mathbf{k}}} \tag{4.133}$$

If $\tilde{\varepsilon}_c$ is within the range

$$w_1 \leq \tilde{\varepsilon}_c \leq w_2 \tag{4.134}$$

then the solution has an important property: there are no localized states in the system. A continuum particle may hop onto the impurity, but after a while it may hop off again. Particles spend only part of their time on the impurity, so it is not a well-defined eigenstate. The impurity state has become a scattering resonance.

Of course, if the new energy $\tilde{\varepsilon}_c$ is outside of the band of continuum states, then a true localized state will exist. Then the solution has a distinct form—a real bound state.

First solve for the case where $\tilde{\varepsilon}_c$ is within the continuous band and no bound state exists. All states are continuum states. A new set of operators $\alpha_{\mathbf{k}}$ and $\alpha_{\mathbf{k}}^\dagger$ are defined which refer to the eigenstates of (4.131). The old operators b and $c_{\mathbf{k}}$ can be expanded in terms of the new set:

$$\begin{aligned} b &= \sum_{\mathbf{k}} \nu_{\mathbf{k}} \alpha_{\mathbf{k}} \\ c_{\mathbf{k}} &= \sum_{\mathbf{k}'} \eta_{\mathbf{k},\mathbf{k}'} \alpha_{\mathbf{k}'} \end{aligned} \tag{4.135}$$

One impurity in the presence of $N \approx 10^{23}$ particles changes their energy by a negligible amount—remember the assumption of no bound states—so that these new operators still have the energy $\varepsilon_{\mathbf{k}}$. All that is needed is the vector $\nu_{\mathbf{k}}$ and the matrix $\eta_{\mathbf{k},\mathbf{k}'}$ for a complete solution. The model is solved first in one dimension. The extension to two and three dimensions will be easy to describe at the end. It is also assumed that the dispersion $\varepsilon_{\mathbf{k}}$ does not permit two states $\mathbf{k}$ and $\mathbf{k}'$ to have the same energy. Of course, this constraint is never realized in practice, since states of $\mathbf{k}$ and $-\mathbf{k}$ usually have the same energy. The problem is that the hopping on and off the impurity preserves energy but not wave vector information. The hopping will mix the states which exist at the same energy. It is a nuisance to describe, so it will be omitted for the

moment, and **k** becomes a scalar. One way around the problem in one dimension is to describe symmetric and anti-symmetric combinations of the states with $\pm k$:

$$c_{ks} = \frac{1}{\sqrt{2}}[c_k + c_{-k}] \tag{4.136}$$

$$c_{ka} = \frac{1}{\sqrt{2}}[c_k - c_{-k}] \tag{4.137}$$

These two states have the same energy ε_k. Only the symmetric state is assumed to interact with the localized state, so that the interaction has the form $\sum_k A_k[b^\dagger c_{ks} + c_{ks}^\dagger b]$. This interpretation of the Hamiltonian makes physical sense. The subscripts s will no longer be written.

Since the new operators α_k and $\alpha_k^\dagger$ describe eigenstates, the Hamiltonian may be written as

$$H = \sum_k \varepsilon_k \alpha_k^\dagger \alpha_k \tag{4.138}$$

The commutator $[b, H]$ is evaluated using both the new operators and the old:

$$[b, H] = b\varepsilon_c + \sum_k A_k c_k = \sum_k \varepsilon_k \nu_k \alpha_k \tag{4.139}$$

Then the old operators are expressed as (4.135), and this equation is

$$\varepsilon_c \sum_k \nu_k \alpha_k + \sum_{kk'} A_k \eta_{k,k'} \alpha_{k'} = \sum_k \varepsilon_k \nu_k \alpha_k \tag{4.140}$$

Each coefficient of α_k is independent; i.e., take the commutator of the above equation with $\alpha_k^\dagger$ and get

$$\nu_k(\varepsilon_k - \varepsilon_c) = \sum_{k'} A_{k'} \eta_{k',k} \tag{4.141}$$

The same procedure is used with the commutator $[c_k, H]$, which gives

$$[c_k, H] = \varepsilon_k c_k + A_k b = \sum_{k'} \eta_{k,k'} \varepsilon_{k'} \alpha_{k'} \tag{4.142}$$

or expressing b and c_k as α_k, gives another equation:

$$\eta_{k,k'}(\varepsilon_k - \varepsilon_{k'}) = -A_k \nu_{k'} \tag{4.143}$$

The last equation gives a result for $\eta_{k,k'}$, when $\varepsilon_k \neq \varepsilon_{k'}$, which means that $k \neq k'$, because of our assumptions. It provides no information about $\eta_{k,k'}$ for the case where $k = k'$. It is necessary to introduce another unknown function Z_k which is proportional to the value of $\eta_{k,k'}$ when $k = k'$:

$$\eta_{k,k'} = -\frac{A_k \nu_{k'}}{\varepsilon_k - \varepsilon_{k'}} + \delta_{kk'} Z_k \nu_k A_k \tag{4.144}$$

The other factors $A_k \nu_k$ are added to the last term for convenience. The energy denominator in the first term is taken as a principal part; the term is omitted when $\varepsilon_k = \varepsilon_{k'}$. All energy denominators are assumed to be principal parts, and the conventional symbol P will be omitted. If the expression for $\eta_{k,k'}$ is used in (4.141), the equation for ν_k becomes

$$\nu_k(\varepsilon_k - \varepsilon_c) = \sum_{k'} A_{k'} \left(-\frac{A_{k'} \nu_k}{\varepsilon_{k'} - \varepsilon_k} + \delta_{kk'} Z_k \nu_k A_k \right) \tag{4.145}$$

The ν_k dependence factors out completely. This factoring leaves an equation in which the only unknown quantity is Z_k,

$$\varepsilon_k - \varepsilon_c = \sum_{k'} \frac{A_{k'}^2}{\varepsilon_k - \varepsilon_{k'}} + Z_k A_k^2 \tag{4.146}$$

so that Z_k quantity is now determined. It also simplifies the notation to introduce the self-energy function

$$\Sigma(\varepsilon_k) = \sum_{k'} \frac{A_{k'}^2}{\varepsilon_k - \varepsilon_{k'}} \tag{4.147}$$

so that Z_k is written as

$$Z_k = \frac{1}{A_k^2}[\varepsilon_k - \varepsilon_c - \Sigma(\varepsilon_k)] \tag{4.148}$$

All quantities on the right-hand side of (4.148) are known, so that Z_k is now known. To make further progress toward obtaining $\eta_{k,k'}$ and ν_k, more equations are needed. They are obtained from the commutation relations for the old operators, which must still be obeyed when they are expressed in terms of the α_k operators::

$$[b, b^\dagger] = 1 = \sum_k \nu_k^2 \tag{4.149}$$

$$[c_k, c_{k'}^\dagger] = \delta_{kk'} = \sum_{k''} \eta_{k,k''}\eta_{k',k''} \tag{4.150}$$

$$[b, c_k^\dagger] = 0 = \sum_{k'} \eta_{k,k'}\nu_{k'} \tag{4.151}$$

The last equation is used first. If the result (4.144) for $\eta_{k,k'}$ is used, then (4.151) becomes

$$0 = \sum_{k'} \nu_{k'}\left(\frac{A_k \nu_{k'}}{\varepsilon_{k'} - \varepsilon_k} + \delta_{kk'} Z_k \nu_k A_k\right) \tag{4.152}$$

or

$$0 = A_k\left(Z_k \nu_k^2 + \sum_{k'} \frac{\nu_{k'}^2}{\varepsilon_{k'} - \varepsilon_k}\right) \tag{4.153}$$

Since A_k is not zero, the quantity in parentheses must vanish. This result will be used below. Next, take Eq. (4.150), insert the result for $\eta_{k,k'}$, and find

$$\delta_{kk'} = \sum_{k''}\left(\frac{A_k \nu_{k''}}{\varepsilon_{k''} - \varepsilon_k} + \delta_{kk''} Z_k \nu_k A_k\right)\left(\frac{A_{k'} \nu_{k''}}{\varepsilon_{k''} - \varepsilon_{k'}} + \delta_{k'k''} Z_{k'} \nu_{k'} A_{k'}\right)$$

which equals

$$\delta_{kk'} = \delta_{kk'} Z_k^2 \nu_k^2 A_k^2 + A_k A_{k'} \frac{Z_k \nu_k^2 - Z_{k'} \nu_{k'}^2}{\varepsilon_k - \varepsilon_{k'}} + A_k A_{k'} \sum_{k''} \frac{\nu_{k''}^2}{(\varepsilon_{k''} - \varepsilon_k)(\varepsilon_{k''} - \varepsilon_{k'})} \tag{4.154}$$

The last term must be rearranged. Since the energy denominators are principal parts, use *Poincaré's theorem*:

$$P\frac{1}{\varepsilon_{k''}-\varepsilon_k}\frac{1}{\varepsilon_{k''}-\varepsilon_{k'}}=P\frac{1}{\varepsilon_k-\varepsilon_{k'}}\left(\frac{1}{\varepsilon_{k''}-\varepsilon_k}-\frac{1}{\varepsilon_{k''}-\varepsilon_{k'}}\right)$$
$$+\pi^2\delta(\varepsilon_{k''}-\varepsilon_k)\delta(\varepsilon_{k''}-\varepsilon_{k'}) \tag{4.155}$$

The delta functions of continuous energy must be changed back to Kronecker deltas, since we are using box normalization in a box of length L. This alteration is

$$\delta(\varepsilon_k-\varepsilon_{k'})=\frac{L}{2\pi v_k}\delta_{kk'} \tag{4.156}$$

where v_k is the velocity of the particle:

$$v_k=\frac{\partial\varepsilon_k}{\partial k} \tag{4.157}$$

The terms can be regrouped in (4.154):

$$\delta_{kk'}=\delta_{kk'}v_k^2A_k^2\left[Z_k^2+\left(\frac{L}{2v_k}\right)^2\right]$$
$$+\frac{A_kA_{k'}}{\varepsilon_k-\varepsilon_{k'}}\left[\left(Z_kv_k^2+\sum_{k''}\frac{v_{k''}^2}{\varepsilon_{k''}-\varepsilon_k}\right)-\left(Z_{k'}v_{k'}^2+\sum_{k''}\frac{v_{k''}^2}{\varepsilon_{k''}-\varepsilon_{k'}}\right)\right] \tag{4.158}$$

The last two quantities in parentheses vanish, since each is identical to (4.153). All the remaining terms are proportional to $\delta_{kk'}$, and so exist only when $k=k'$. These steps give the equation for v_k:

$$v_k^2=\left\{A_k^2\left[Z_k^2+\left(\frac{L}{2v_k}\right)^2\right]\right\}^{-1} \tag{4.159}$$

Of course, once v_k is known, then $\eta_{k,k'}$ is obtained easily from (4.144). The result for v_k may be reworked into a form in which the physics is more transparent. The earlier result (4.148) for Z_k is used to rewrite (4.159) as

$$v_k^2=\frac{A_k^2}{[\varepsilon_k-\varepsilon_c-\Sigma(\varepsilon_k)]^2+(LA_k^2/2v_k)^2} \tag{4.160}$$

Furthermore, the self energy $\Sigma(\varepsilon_k)$ will be interpreted as the real part of the retarded self-energy:

$$\Sigma_{\text{ret}}(\varepsilon)=\sum_{k'}\frac{A_{k'}^2}{\varepsilon-\varepsilon_{k'}+i\delta} \tag{4.161}$$

$$\text{Re}[\Sigma_{\text{ret}}(\varepsilon_k)]=P\sum_{k'}\frac{A_{k'}^2}{\varepsilon_k-\varepsilon_{k'}}=\Sigma(\varepsilon_k) \tag{4.162}$$

Similarly, the imaginary part of the retarded self-energy is

$$\text{Im}[\Sigma_{\text{ret}}(\varepsilon)]=-\pi\sum_{k'}A_{k'}^2\delta(\varepsilon-\varepsilon_{k'}) \tag{4.163}$$

By using the previous identity (4.156) for the delta function, the imaginary part of the retarded self-energy becomes

$$\mathrm{Im}[\Sigma_{\mathrm{ret}}(\varepsilon)] = -\pi \sum_{k'} A_{k'}^2 \left(\frac{L}{2v_k}\right) \delta_{kk'} = -\frac{L}{2v_k} A_k^2 \tag{4.164}$$

The expression on the right is recognized as the same factor which occurs in the denominator of (4.160) for ν_k^2. In fact, this quantity may be rewritten as

$$\nu_k^2 = -\left(\frac{2v_k}{L}\right) \frac{\mathrm{Im}[\Sigma(\varepsilon_k)]}{[\varepsilon_k - \varepsilon_c - \mathrm{Re}\ \Sigma(\varepsilon_k)]^2 + [\mathrm{Im}\ \Sigma(\varepsilon_k)]^2} \tag{4.165}$$

$$= -\left(\frac{2v_k}{L}\right) \mathrm{Im}\left\{\frac{1}{\varepsilon_k - \varepsilon_c - \Sigma_{\mathrm{ret}}(\varepsilon_k)}\right\} \tag{4.166}$$

so that ν_k^2 is proportional to the imaginary part of a retarded Green's function. It was shown in Sec. 3.3 that the imaginary part of a retarded Green's function is proportional to the spectral function:

$$A(\varepsilon) = -2\,\mathrm{Im}\left\{\frac{1}{\varepsilon - \varepsilon_c - \Sigma(\varepsilon)}\right\} \tag{4.167}$$

$$\nu_k^2 = \frac{v_k}{L} A(\varepsilon_k) \tag{4.168}$$

The remaining equation (4.149) which must be satisfied may now be considered:

$$1 = \sum_k \nu_k^2 \tag{4.169}$$

Since ν_k^2 is now known, it is readily substituted, and the integral is:

$$1 = \sum_k \frac{v_k}{L} A(\varepsilon_k) = \int \frac{d\varepsilon_k}{2\pi} A(\varepsilon_k) \tag{4.170}$$

In the second equality the integration variable was altered to the energy of the particle. This integral is just the sum rule for spectral functions which was proved in (3.121). The sum rule applies to any spectral function and must apply to the one derived above. Equation (4.149) is satisfied. The transformation (4.135) to the α_k has been achieved. The form derived for ν_k satisfies all the commutation relations (4.149)–(4.151) as well as the commutation relations with the original Hamiltonian. The problem has been solved exactly.

Since the quantity

$$G(\varepsilon) = \frac{1}{\varepsilon - \varepsilon_c - \Sigma_{\mathrm{ret}}(\varepsilon)} \tag{4.171}$$

is identified as a retarded Green's function, for which particle is it the Green's function? It is the Green's function of the localized particle, or what became of the localized particle after it

became delocalized. This assertion is somewhat evident by considering the definition of the localized retarded Green's function for fermions:

$$\begin{aligned} G(t) &= -i\Theta(t)\langle b(t)b^\dagger(0) + b^\dagger(0)b(t)\rangle \\ &= -i\Theta(t)\sum_k v_k^2\langle \alpha_k(t)\alpha_k^\dagger(0) + \alpha_k^\dagger(0)\alpha_k(t)\rangle \\ &= -i\Theta(t)\sum_k v_k^2 e^{-i\varepsilon_k t} = \frac{-i}{L}\Theta(t)\sum_k v_k A(\varepsilon_k)e^{-i\varepsilon_k t} \\ &= -i\Theta(t)\int_{-\infty}^{\infty}\frac{d\varepsilon_k}{2\pi}A(\varepsilon_k)e^{-i\varepsilon_k t} \end{aligned} \tag{4.172}$$

In the last step the summation is changed to an integral over ε_k. The last line is the identity for the retarded Green's function assigned in Problem 6 of Chapter 3.

The Fano–Anderson model is just a description of a localized scattering resonance. The continuum particles come to the impurity, spend some time in the resonant state, and then depart in another continuum state. This model is really no different from that of the prior section, on impurity scattering, if the impurity potential is made to have a resonance. The phase shifts for the present model are defined as

$$\tan[\delta(\varepsilon)] = \frac{\mathrm{Im}[\Sigma_{\mathrm{ret}}(\varepsilon)]}{\varepsilon - \varepsilon_c - \mathrm{Re}\,\Sigma_{\mathrm{ret}}(\varepsilon)} \tag{4.173}$$

The phase shift is all that is needed to use the results of the prior section to describe the Fano–Anderson model. The resonant behavior comes from approximating $-\mathrm{Im}\,\Sigma \approx \Gamma$ as a constant, or at least as a slowly varying function of energy. Then the spectral function $A(\varepsilon)$ is Lorentzian. If the width Γ is small, this function describes a resonance.

As an example of a scattering potential, consider the potential $V(r)$ shown in Fig. 4.4(a). The resonance behavior is chosen by first solving for the bound states of Fig. 4.4(b). For the latter case, bound states exist when

$$\tan(ka) = -\frac{k}{\alpha} \tag{4.174}$$

$$\alpha = \left(\frac{2mV_0}{\hbar^2} - k^2\right)^{1/2} \tag{4.175}$$

A bound state at half of the well depth is made by choosing $ka = \alpha a = 3\pi/4$. Here k^2 is the particle energy relative to the bottom of the well. The well shape of Fig. 4.4(a) will have a resonance at nearly the same energy. The finite extent of the repulsive part will allow the particle to leave and impart a width to the state and to the scattering resonance. The phase shift is shown in Fig. 4.4(c) for the potential of part (a). The steep rise in the phase shift, of about π, occurs at the value $ka \approx 3\pi/4$.

The change in phase shift by π across the resonance may be understood from the Friedel sum rule. As the eigenstates of the system are filled up, one particle will reside in the vicinity of the impurity and contribute unity to the Friedel sum over δ/π.

The initial Hamiltonian was written as (4.131) in terms of eigenstates which were not orthogonal. All of our effort was merely an exercise in orthogonalization. This work could have been avoided by writing H in an orthogonal basis. The basis would have the particles with a scattering resonance from the impurity. The Hamiltonian describes one-particle behavior.

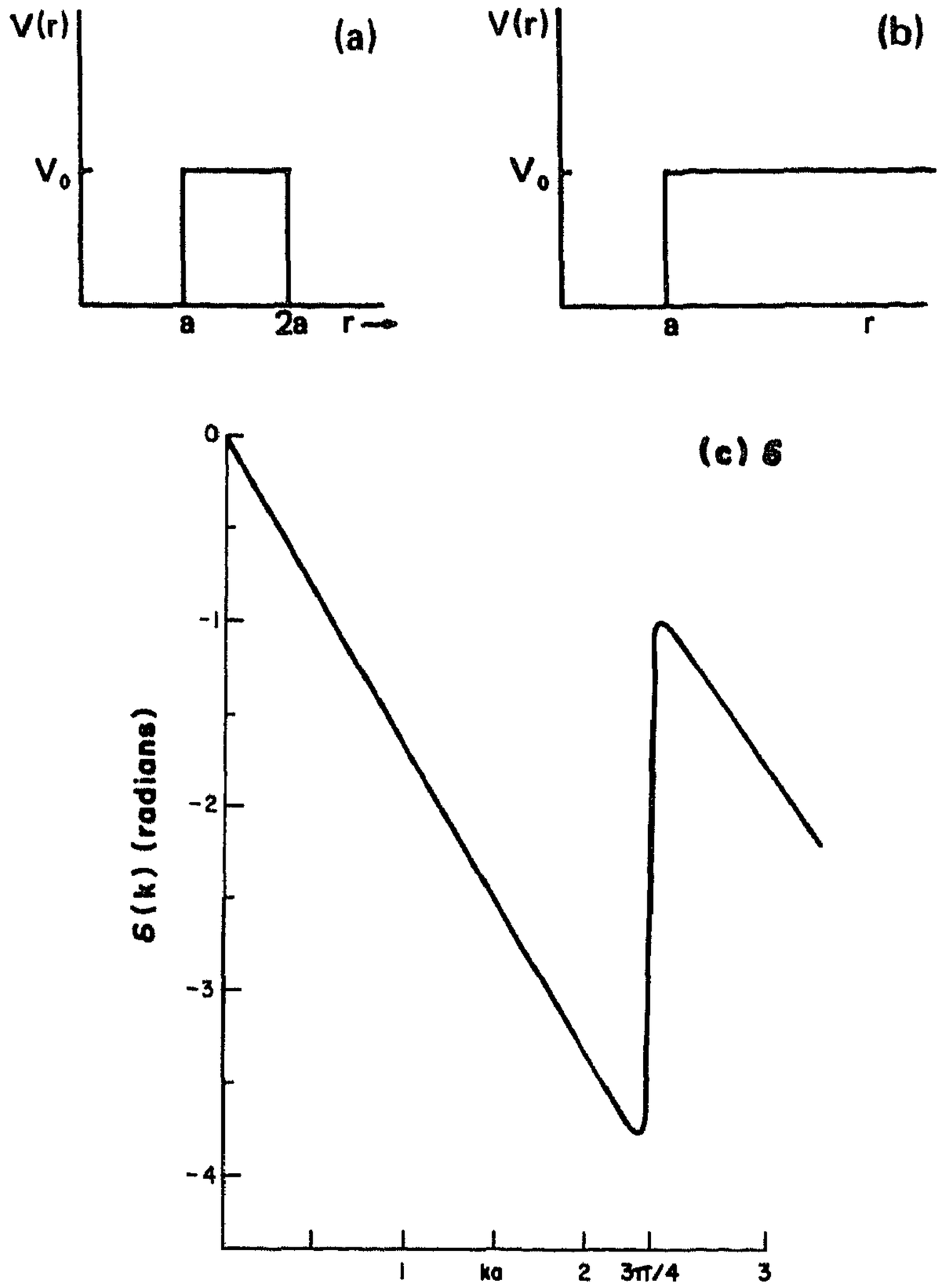


FIGURE 4.4 (a) A potential whose phase shifts can show resonance behavior. The resonance occurs at the bound state of the potential shown in (b). (c) The phase shift jumps in value by almost π radians at the wave vector of the resonance.

The model is very important in a number of applications. The *d*-states of transition metals seem well described as scattering resonances of the *sp* electrons (see Anderson and McMillan, 1967). Another example is in surface physics. When atoms come from a gas and are absorbed on the surface of a metal, the conduction electrons of the metal may hop onto the atomic-like states of the absorbed atom. The charging effect on the absorbed atoms seems well described by the model (Schönhammer and Gunnarsson, 1977). It has also found numerous applications in atomic physics.

The model will now be solved by using Green's functions. This solution is much quicker than the prior method, since there is only one self-energy diagram. The Hamiltonian is written as $H_0 + V$, where V is the last term in (4.131) and H_0 is the first two terms. The Green's function for the localized state is:

$$\mathcal{G}(ip) = -\int_0^\beta d\tau e^{ip\tau} \langle T_\tau b(\tau) b^\dagger(0) \rangle \tag{4.176}$$

The first self-energy term comes from the $n = 2$ term in the S-matrix expansion:

$$\begin{aligned} &-\frac{1}{2}\int_0^\beta d\tau e^{ip\tau} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \sum_{k_1 k_2} A_{k_1} A_{k_2} \langle T_\tau b(\tau) \\ &\times [c_{k_1}(\tau_1) b^\dagger(\tau_1) + b(\tau_1) c_{k_1}^\dagger(\tau_1)][c_{k_2}(\tau_2) b^\dagger(\tau_2) + b(\tau_2) c_{k_2}^\dagger(\tau_2)] b^\dagger(0) \rangle \end{aligned} \tag{4.177}$$

The correlation function is easily evaluated in terms of the unperturbed Green's functions of the localized state $\mathscr{G}^{(0)}(\tau)$ and the continuum states $\mathscr{G}^{(0)}(k, \tau)$ to give

$$\sum_k A_k^2 \int_0^\beta d\tau e^{ip\tau} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \mathscr{G}^{(0)}(\tau - \tau_1)\mathscr{G}^{(0)}(k, \tau_1 - \tau_2)\mathscr{G}^{(0)}(\tau_2)$$
$$= \sum_k A_k^2 \mathscr{G}^{(0)}(ip)^2 \mathscr{G}^{(0)}(k, ip) = \mathscr{G}^{(0)}(ip)^2 \Sigma(ip) \tag{4.178}$$

$$\Sigma(ip) = \sum_k A_k^2 \mathscr{G}^{(0)}(k, ip) = \sum_k \frac{A_k^2}{ip - \varepsilon_k} \tag{4.179}$$

$\Sigma(ip)$ turns out to be the only self-energy diagram. The higher terms in the S matrix only produce higher powers of this self-energy contribution:

$$\mathscr{G}(ip) = \mathscr{G}^{(0)}(ip)[1 + \mathscr{G}^{(0)}\Sigma + (\mathscr{G}^{(0)}\Sigma)^2 + \cdots] \tag{4.180}$$

The series may be summed to give the Dyson equation for the Matsubara Green's function:

$$\mathscr{G}(ip) = \frac{1}{ip - \varepsilon_c - \Sigma(ip)} \tag{4.181}$$

Changing $ip \to \varepsilon - \mu + i\delta$ gives the retarded Green's function, which is the same as noted earlier in (4.171). This derivation proves that v_k^2 is really proportional to the spectral function of the $G(\varepsilon)$. The equivalence of this result to (4.172) is just an example of the general theorem proved in Problem 6 of Chapter 3:

$$G_{\text{ret}}(t) = -i\Theta(t) \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} A(\varepsilon) e^{-i\varepsilon t} \tag{4.182}$$

These results may be generalized to higher dimension and other energy bands. This extension is most easily accomplished using the Green's function technique, since the steps are the same as in one dimension. The Matsubara form of the Green's function of the b operators (4.172) is still given by Eq. (4.181) where the self-energy operator is summed over all states in the system:

$$\Sigma(ip) = \sum_{\mathbf{k}} \frac{A_{\mathbf{k}}^2}{ip - \varepsilon_{\mathbf{k}}} \tag{4.183}$$

The real part of the denominator of the retarded Green's function is ($ip \to \varepsilon$)

$$\varepsilon - \varepsilon_c - \text{Re}[\Sigma_{\text{ret}}(\varepsilon)] \tag{4.184}$$

The Green's function has poles at the points where the denominator vanishes. The poles of the Green's function corresponds to excitations of the system. The poles are at the energy $\tilde{\varepsilon}_c$ which satisfies

$$\tilde{\varepsilon}_c = \varepsilon_c + \text{Re}[\Sigma_{\text{ret}}(\tilde{\varepsilon}_c)] \tag{4.185}$$

The above expression for $\tilde{\varepsilon}_c$ is the result asserted earlier in (4.133). If the pole lies within the continuum of states, then the resonance occurs, and there is no bound state. But if the pole occurs outside of the band of continuum states, then the system has a real bound state. A pole occurs when $\text{Im}\,\Sigma = 0$ at the same point that the real part of the denominator vanishes. Generally, $\text{Im}\,\Sigma$ is not zero throughout the continuum band, so that $\text{Im}\,\Sigma = 0$ only outside of the band.

If it is assumed that the bound state occurs outside of the band, the spectral function has the form

$$A(\varepsilon) = 2\pi\delta(\varepsilon - \varepsilon_c - \mathrm{Re}[\Sigma(\varepsilon)]) + \frac{-2\ \mathrm{Im}\ \Sigma(\varepsilon)}{[\varepsilon - \varepsilon_c - \mathrm{Re}\ \Sigma]^2 + [\mathrm{Im}\ \Sigma]^2} \tag{4.186}$$

The first term comes from the pole of the Green's function, while the second term is from those parts where $\mathrm{Im}\ \Sigma \neq 0$. The first term may be rewritten as a simple delta function plus a renormalization factor. This procedure was described earlier in (3.155):

$$A(\varepsilon) = 2\pi Z\delta(\varepsilon - \tilde{\varepsilon}_c) + \frac{-2\ \mathrm{Im}\ \Sigma}{[\varepsilon - \varepsilon_c - \mathrm{Re}\ \Sigma]^2 + [\mathrm{Im}\ \Sigma]^2} \tag{4.187}$$

$$Z = \left(1 - \frac{d\ \mathrm{Re}\ \Sigma}{d\varepsilon}\right)^{-1}_{\tilde{\varepsilon}_c} \tag{4.188}$$

In other examples there may be more than one bound state, in which case the first term becomes a series of delta functions.

These points are well illustrated by an example. Take a one-dimensional tight-binding model on a solid of unit separation $a = 1$ and length L with a constant coupling constant:

$$\varepsilon_k = -2w\ \cos(k) \tag{4.189}$$

$$A_k^2 = \frac{C}{L} \tag{4.190}$$

Also, take the initial bound state energy ε_c in the middle of the band $\varepsilon_c = 0$. The band structure is shown in Fig. 4.5. The Brillouin zone is defined by $-\pi \leq k \leq \pi$. The self-energy function $\Sigma_{\mathrm{ret}}(\varepsilon)$ is now elementary to evaluate:

$$\Sigma_{\mathrm{ret}}(\varepsilon) = \frac{C}{L}\sum_k \frac{1}{\varepsilon + w\ \cos(k) + i\delta} = \frac{C}{2\pi}\int_{-\pi}^{\pi} \frac{dk}{\varepsilon + w\ \cos(k) + i\delta}$$

The real part is

$$\begin{aligned} \mathrm{Re}[\Sigma] &= \frac{C}{2\pi}\int_{-\pi}^{\pi} \frac{dk}{\varepsilon + w\ \cos(k)} = \frac{C\ \mathrm{sgn}(\varepsilon)}{\sqrt{\varepsilon^2 - w^2}}, \quad \text{if } \varepsilon^2 > w^2 \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \varepsilon^2 < w^2 \end{aligned} \tag{4.191}$$

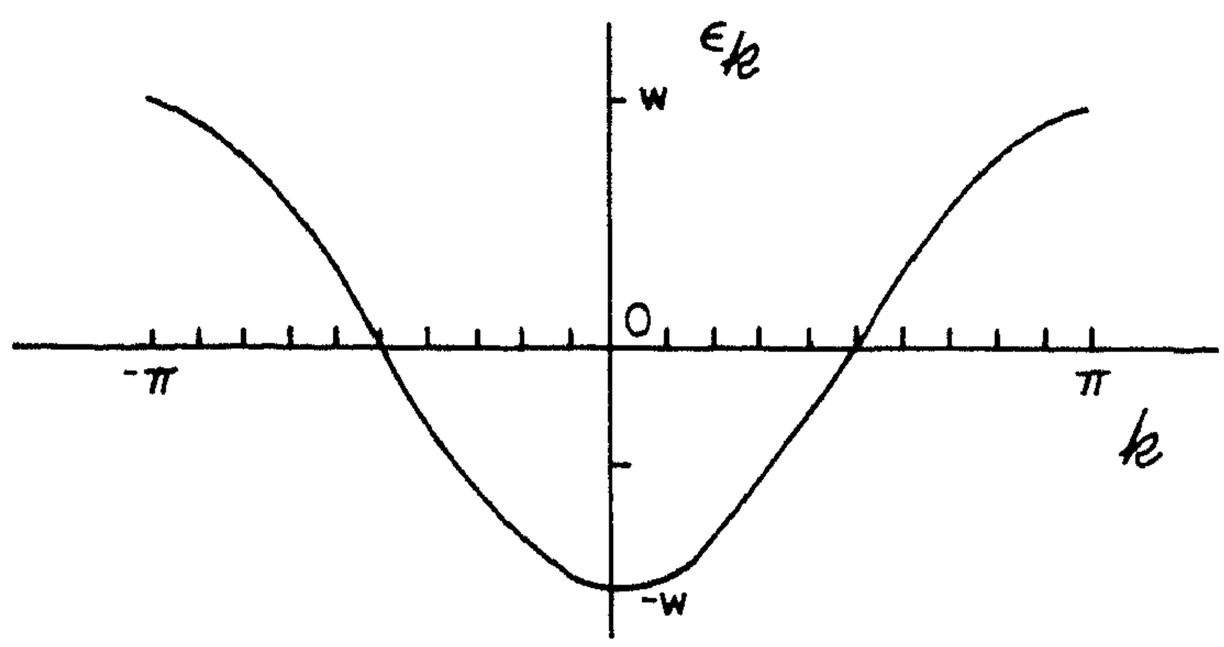


FIGURE 4.5 Tight-binding energy bands in one dimension.

where sgn(ε) is the sign of ε. The imaginary part is

$$-\mathrm{Im}[\Sigma(\varepsilon)] = \frac{\pi C}{2\pi}\int_{-\pi}^{\pi} \delta[\varepsilon + w\ \cos(k)] \tag{4.192}$$

$$= \frac{C}{w|\sin(k)|} = \frac{C}{\sqrt{w^2 - \varepsilon^2}} \text{ if } w^2 > \varepsilon^2 \tag{4.193}$$

The spectral function of the localized state is

$$A(\varepsilon) = \Theta(\varepsilon^2 - w^2)\left\{2\pi\delta\left[\varepsilon - \frac{C}{\sqrt{\varepsilon^2 - w^2}}\right] + 2\pi\delta\left[\varepsilon + \frac{C}{\sqrt{\varepsilon^2 - w^2}}\right]\right\}$$
$$+ \frac{2C\Theta(w^2 - \varepsilon^2)}{\sqrt{w^2 - \varepsilon^2}[\varepsilon^2 + C^2/(w^2 - \varepsilon^2)]} \tag{4.194}$$

The step function $\Theta(x)$ is unity if the argument is positive and zero if negative. The first term is nonzero only outside of the band of continuum states, while the last term is nonzero only inside the band. Figure 4.6 shows a plot of the entire spectral function. There are two sharp bound states, one below the band and one above. The continuous contribution throughout the band is due to the last term in (4.194). According to the sum rule, the total area under all the contributions must be 2π. The solution to the equation

$$\tilde{\varepsilon}_c = \frac{C}{\sqrt{\tilde{\varepsilon}_c^2 - w^2}} \tag{4.195}$$

is

$$\tilde{\varepsilon}_c = \pm\frac{1}{\sqrt{2}}[w^2 + \sqrt{w^4 + 4C^2}]^{1/2} \tag{4.196}$$

and the renormalization factor is

$$Z_c = \frac{\tilde{\varepsilon}_c^2 - w^2}{2\tilde{\varepsilon}_c^2 - w^2} \tag{4.197}$$

and is the same at both poles. For strong coupling, where C becomes very large, $\tilde{\varepsilon}_c \gg w$. The renormalization factors approach $\frac{1}{2}$, so that one-half the spectral weight is in each pole.

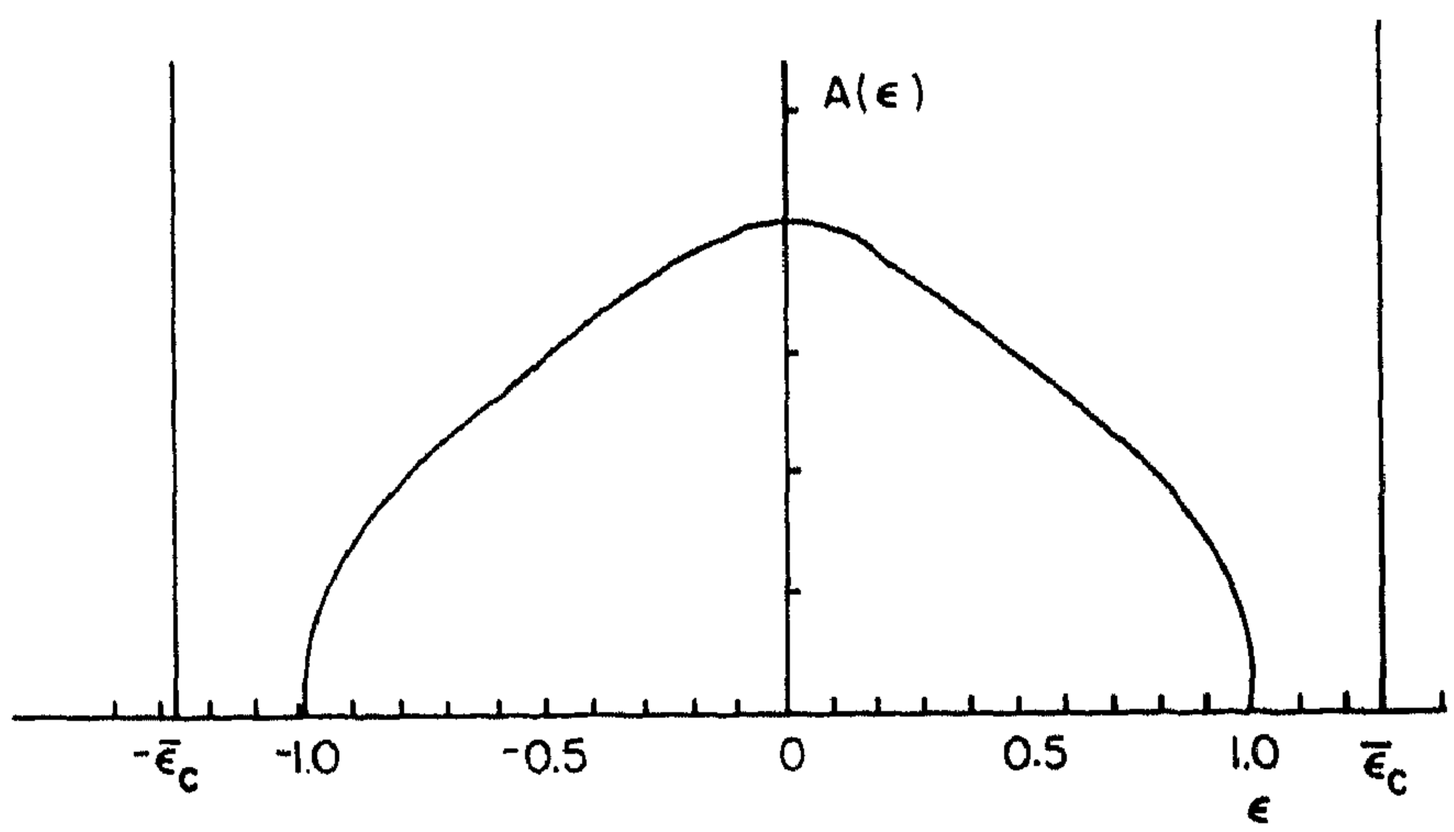


FIGURE 4.6 Spectral function from Eqn. (4.196). $C = 1, \omega = 1.0$.

4.3. INDEPENDENT BOSON MODELS

The independent boson model is very important in many-body physics. It is an exactly solvable model which describes some relaxation phenomena. It has become very useful for describing a wide variety of effects in solid state physics. This section discusses it at great length. Two different derivations of the basic mathematical result are provided. In addition, several variations on the model are described briefly. An exact solution may be obtained by a variety of techniques. The usual procedure is followed: first solve it by ordinary operator algebra, and later solve it using Green's functions.

The first Hamiltonian which will be solved is

$$H = c^\dagger c\left[\varepsilon_c + \sum_{\mathbf{q}}(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger)\right] + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \tag{4.198}$$

The Hamiltonian describes a fixed particle of energy ε_c interacting with a set of phonons with frequencies $\omega_{\mathbf{q}}$. The interaction occurs only when the state is occupied and $c^\dagger c = 1$. The phonons are the independent bosons. An alert reader will note that this model was solved exactly in Chapter 1. By making a canonical transformation, the Hamiltonian may be rewritten as [see (1.239)]

$$\bar{H} = c^\dagger c(\varepsilon_c - \Delta) + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \tag{4.199}$$

where the self-energy is

$$\Delta = \sum_{\mathbf{q}} \frac{M_{\mathbf{q}}^2}{\omega_{\mathbf{q}}} \tag{4.200}$$

The solution to this problem is identical with the problem of a charge on a harmonic spring in a uniform electric field. The electric field causes a displacement of the charge to a new equilibrium position, about which it vibrates with the same frequency as before; see (1.38).

The present objective is to obtain a better description of the fluctuations about equilibrium. The self-energy Δ is for the zero temperature ground state configuration of the system. To study relaxation effects, the fluctuations need to be understood. They are obtained from evaluating the time variation of the Green's function.

$$G(t) = -i\langle T_t c(t) c^\dagger(0)\rangle \tag{4.201}$$

It is solved for the real-time Green's function, which is permissible in the present case because the single impurity state $c^\dagger c$ will not alter the phonon energies. In the thermodynamic averaging over the phonon states, the perturbation V has no effect, so there is no need to worry about the perturbation expansion for the factor $\exp(-\beta H)$. First solve the Green's function of time at nonzero temperature.

4.3.1. Solution by Canonical Transformation

The Hamiltonian is first solved by a canonical transformation. This solution is nearly identical to the procedure described in Chapter 1, but here it is done with more rigor. A new Hamiltonian is desired by a transformation of the type

$$\bar{H} = e^S H e^{-S} = c^\dagger c(\varepsilon_c - \Delta) + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \tag{4.202}$$

The transformation must be done so that $S^\dagger = -S$. The transformation on any product of operators is done by taking the product of the transformed operators. The last theorem is shown by inserting $1 = e^{-S}e^{S}$ between each operator:

$$e^{S}ABCD\cdots Ze^{-S} = (e^{S}Ae^{-S})(e^{S}Be^{-S})(e^{S}Ce^{-S})\cdots(e^{S}Ze^{-S}) = \bar{A}\bar{B}\bar{C}\cdots\bar{Z}$$

It is assumed that any function of operators may be expressed as a power series, and the transformation on a function of operators is just the function of the transformed operators:

$$e^{S}f(A)e^{-S} = e^{S}\sum_{n=0}^{\infty} a_n A^n e^{-S} = \sum_{n=0}^{\infty} a_n(\bar{A})^n = f(\bar{A}) \tag{4.203}$$

The transformation is considered on each operator separately, and the transformed Hamiltonian is the old one with the new operators. These are evaluated using

$$\bar{A} = e^{S}Ae^{-S} = A + [S, A] + \frac{1}{2!}[S, [S, A]] + \cdots \tag{4.204}$$

$$S = c^\dagger c\sum_{\mathbf{q}}\frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}}) \tag{4.205}$$

which gives

$$\begin{aligned}
\bar{c} &= cX \\
\bar{c}^\dagger &= c^\dagger X^\dagger \\
\bar{a}_{\mathbf{q}} &= a_{\mathbf{q}} - \frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}c^\dagger c \\
\bar{a}_{\mathbf{q}}^\dagger &= a_{\mathbf{q}}^\dagger - \frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}c^\dagger c
\end{aligned} \tag{4.206}$$

where the operator is

$$X = \exp\left[-\sum_{\mathbf{q}}\frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})\right] \tag{4.207}$$

Since X commutes with the c operator, the number operator is the same in the new representation,

$$\bar{c}^\dagger\bar{c} = c^\dagger cX^\dagger X = c^\dagger c \tag{4.208}$$

since $X^\dagger = X^{-1}$. The transformed Hamiltonian is

$$\begin{aligned}
\bar{H} &= \varepsilon_c\bar{c}^\dagger\bar{c} + \sum_{\mathbf{q}}\omega_{\mathbf{q}}\left(a_{\mathbf{q}}^\dagger - \frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}c^\dagger c\right)\left(a_{\mathbf{q}} - \frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}c^\dagger c\right) \\
&\quad + \sum_{\mathbf{q}}M_{\mathbf{q}}\left(a_{\mathbf{q}}^\dagger + a_{\mathbf{q}} - 2\frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}c^\dagger c\right)c^\dagger c \qquad (4.209) \\
&= c^\dagger c(\varepsilon_c - \Delta) + \sum_{\mathbf{q}}\omega_{\mathbf{q}}a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \qquad (4.210)
\end{aligned}$$

This transformed Hamiltonian is precisely the form which was the objective in (4.202). It is also the form which was derived earlier in Chapter 1. The reason for repeating the analysis is that the important factors X in (4.207) are needed for the discussion of fluctuations.

This transformation is now applied to the Green's function (4.201). The factor $1 = e^{-S}e^{S}$ is inserted into the trace, say for $t > 0$:

$$\mathrm{Tr}(e^{-\beta H}e^{iHt}ce^{-iHt}c^{\dagger}e^{-S}e^{S}) \tag{4.211}$$

Using the cyclic properties of the trace,

$$\mathrm{Tr}(e^{S}e^{-\beta H}e^{iHt}ce^{-iHt}c^{\dagger}e^{-S}) = \mathrm{Tr}(e^{-\beta \bar{H}}e^{i\bar{H}t}\bar{c}e^{-i\bar{H}t}\bar{c}^{\dagger}) \tag{4.212}$$

By using the previous theorems, everything in the trace is now changed to the transformed representation. The Green's function may be written as ($t > 0$)

$$G(t) = -ie^{\beta\Omega}\mathrm{Tr}(e^{-\beta\bar{H}}e^{i\bar{H}t}\bar{c}e^{-i\bar{H}t}\bar{c}^{\dagger}) \tag{4.213}$$

It should be emphasized that this $G(t)$ will be exactly equal to the earlier definition (4.201). The new equation for $G(t)$ is just another way of evaluating the same thing.

At first glance it appears that this evaluation is now trivial, since the Hamiltonian is diagonal in the $\bar{c}^{\dagger}\bar{c}$-operators. But this conclusion is untrue and misleading, since $\bar{c}$ and $\bar{c}^{\dagger}$ do not commute with a or $a^{\dagger}$ because of the X factor in (4.206). It is necessary to stick with the c and a representation and to put in X explicitly. The Green's function becomes

$$G(t) = -ie^{\beta\Omega}\mathrm{Tr}(e^{-\beta\bar{H}}e^{i\bar{H}t}cXe^{-i\bar{H}t}c^{\dagger}X^{\dagger}) \tag{4.214}$$

However, it is now possible to achieve a great simplification. Since $\bar{H}$ is diagonal in c and a, it is easy to commute it through cX and to obtain the time development of the operators:

$$e^{i\bar{H}t}cXe^{-i\bar{H}t} = e^{-it(\varepsilon_c - \Delta)}cX(t) \tag{4.215}$$

$$X(t) = \exp\left[-\sum_{\mathbf{q}}\frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}(a_{\mathbf{q}}^{\dagger}e^{i\omega_{\mathbf{q}}t} - a_{\mathbf{q}}e^{-i\omega_{\mathbf{q}}t})\right] \tag{4.216}$$

The phonon and electron parts of the trace may now be completely separated. For ($t > 0$):

$$G(t) = -ie^{\beta\Omega}\mathrm{Tr}(e^{-\beta\bar{\varepsilon}_c n}cc^{\dagger})\mathrm{Tr}[e^{-\beta H_p}X(t)X^{\dagger}(0)] \tag{4.217}$$

$$\bar{\varepsilon}_c = \varepsilon_c - \Delta \tag{4.218}$$

$$H_p = \sum_{\mathbf{q}}\omega_{\mathbf{q}}a_{\mathbf{q}}^{\dagger}a_{\mathbf{q}} \tag{4.219}$$

The result for $G(t)$ has the great simplification mentioned above. The particle part is trivial (assuming they are fermions),

$$e^{\beta\Omega_{el}}\mathrm{Tr}(e^{-\beta\bar{\varepsilon}_c n}cc^{\dagger}) = 1 - n_F(\bar{\varepsilon}_c) \tag{4.220}$$

and there remains only the problem of evaluating the phonon part of the trace. This evaluation is nontrivial, although it may be done exactly. The operator disentangling method was introduced by Feynman (1951).

4.3.2. Feynman Disentangling of Operators

The objective is now to evaluate the trace over the phonon distributions of the operator:

$$F(t) = e^{\beta\Omega_{ph}}\mathrm{Tr}[e^{-\beta\sum_{\mathbf{q}}\omega_{\mathbf{q}}n_{\mathbf{q}}}X(t)X^{\dagger}(0)] \tag{4.221}$$

Each wave vector state $\mathbf{q}$ is averaged independently, and the final result is the product over $\mathbf{q}$ states:

$$\begin{aligned} F(t) &= \Pi_{\mathbf{q}}\mathscr{F}_{\mathbf{q}}(t) \\ \mathscr{F}_{\mathbf{q}} &= e^{\beta\Omega_{\mathbf{q}}}\sum_{n_{\mathbf{q}}=0}^{\infty} e^{-\beta n_{\mathbf{q}}\omega_{\mathbf{q}}}\langle n_{\mathbf{q}}|e^{-\lambda_{\mathbf{q}}B_{\mathbf{q}}(t)}e^{\lambda_{\mathbf{q}}B_{\mathbf{q}}(0)}|n_{\mathbf{q}}\rangle \\ \lambda_{\mathbf{q}} &= \frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}} \\ B_{\mathbf{q}}(t) &= a_{\mathbf{q}}^{\dagger}e^{i\omega\mathbf{q}t} - a_{\mathbf{q}}e^{-i\omega\mathbf{q}t} \end{aligned} \tag{4.222}$$

For each state $\mathbf{q}$, the trace is merely a summation over all possible integer values of $n_{\mathbf{q}}$ between zero and infinity. The prefactor is the normalization:

$$e^{\beta\Omega_{\mathbf{q}}} = \left(\sum_{n_{\mathbf{q}}=0}^{\infty} e^{-\beta n_{\mathbf{q}}\omega_{\mathbf{q}}}\right)^{-1} = 1 - e^{-\beta\omega_{\mathbf{q}}} \tag{4.223}$$

The present notation is simplified by first dropping all $\mathbf{q}$ subscripts

$$\mathscr{F}(t) = (1 - e^{-\beta\omega})\sum_{n=0}^{\infty} e^{-\beta n\omega}\langle n|e^{-\lambda(a^{\dagger}e^{i\omega t} - ae^{-i\omega t})}e^{\lambda(a^{\dagger}-a)}|n\rangle \tag{4.224}$$

The state $|n\rangle$ is the one with n excitations and is given in terms of the operators as

$$|n\rangle = \frac{(a^{\dagger})^n}{\sqrt{n!}}|0\rangle \tag{4.225}$$

The first step is to separate the operators in the exponential. Use Feynman's theorem on the disentangling of operators, which is as follows:

Theorem: If the operators A and B have the property that their commutator $C = [A, B]$ commutes with both A and B, then

$$e^{A+B} = e^{A}e^{B}e^{-1/2[A,B]} \tag{4.226}$$

This theorem was proved in Problem 5 of Chapter 2. It is used to separate the exponents in $X(t)$ and $X^{\dagger}(0)$. To evaluate $X(t)$, set

$$A = -\lambda a^{\dagger}e^{i\omega t} \tag{4.227}$$

$$B = \lambda a e^{-i\omega t} \tag{4.228}$$

$$[A, B] = -\lambda^2[a^{\dagger}, a] = \lambda^2 \tag{4.229}$$

and

$$X(t) = e^{A+B} = e^{-\lambda^2/2} e^{-\lambda a^\dagger e^{i\omega t}} e^{\lambda a e^{-i\omega t}} \tag{4.230}$$

$$X^\dagger(0) = e^{-\lambda^2/2} e^{\lambda a^\dagger} e^{-\lambda a} \tag{4.231}$$

$$X(t)X^\dagger(0) = e^{-\lambda^2} e^{-\lambda a^\dagger e^{i\omega t}} e^{\lambda a e^{-i\omega t}} e^{\lambda a^\dagger} e^{-\lambda a} \tag{4.232}$$

The result for $X^\dagger(0)$ is just the Hermitian conjugate of $X(t)$ evaluated at $t = 0$. The factor of $X(t)X^\dagger(0)$ is what is needed in the average in (4.224).

The next step is to get all the destruction operators on the right and the creation operators on the left. The center two operators need to be exchanged. Since they do not commute, this exchange will produce another complex phase factor. These two operators are written as

$$e^{\lambda a(t)} e^{\lambda a^\dagger} = e^{\lambda a^\dagger} [e^{-\lambda a^\dagger} e^{\lambda a(t)} e^{\lambda a^\dagger}] \tag{4.233}$$

The factor in brackets has exactly the form derived earlier in (4.203) where $S = -\lambda a^\dagger$. Consider

$$e^{-\lambda a^\dagger} a e^{\lambda a^\dagger} = a - \lambda[a^\dagger, a] + \frac{\lambda^2}{2!}[a^\dagger[a^\dagger, a]] \ldots \tag{4.234}$$

$$= a + \lambda \tag{4.235}$$

then

$$e^{-\lambda a^\dagger} e^{\lambda a(t)} e^{\lambda a^\dagger} = \exp[\lambda e^{-i\omega t}(a + \lambda)] = e^{\lambda^2 e^{-i\omega t}} e^{\lambda a(t)} \tag{4.236}$$

so (4.224) is finally arranged into the desired form:

$$\mathscr{F}(t) = (1 - e^{-\beta\omega}) e^{-\lambda^2(1-e^{-i\omega t})} \tag{4.237}$$

$$\times \sum_{n=0}^{\infty} e^{-\beta\omega n} \langle n | e^{\lambda a^\dagger (1-e^{i\omega t})} e^{-\lambda a(1-e^{-i\omega t})} | n \rangle \tag{4.238}$$

All the terms with a can be collected together in one exponential since all of these terms commute. Likewise for all the terms with $a^\dagger$. Next prove that

$$(1 - e^{-\beta\omega}) \sum_{n=0}^{\infty} e^{-\beta n\omega} \langle n | e^{u^* a^\dagger} e^{-ua} | n \rangle = e^{-|u|^2 N} \tag{4.239}$$

$$N = \frac{1}{e^{\beta\omega} - 1} \tag{4.240}$$

where, for our case, $u = \lambda(1 - e^{-i\omega t})$. Equation (4.240) is proved by expanding the exponents in a power series:

$$e^{-ua} | n \rangle = \sum_{l=0}^{\infty} \frac{(-u)^l}{l!} a^l | n \rangle \tag{4.241}$$

Now recall the properties of boson destruction operators acting on a harmonic oscillator state:

$$a|n\rangle = n^{1/2}|n-1\rangle \tag{4.242}$$

$$a^2|n\rangle = [n(n-1)]^{1/2}|n-2\rangle \tag{4.243}$$

$$a^l|n\rangle = \left[\frac{n!}{(n-l)!}\right]^{1/2}|n-l\rangle \tag{4.244}$$

The feature that $a^l|n\rangle = 0$ for $l > n$ is very useful: It terminates the power series after n terms:

$$e^{-ua}|n\rangle = \sum_{l=0}^{n}\frac{(-u)^l}{l!}\left[\frac{n!}{(n-l)!}\right]^{1/2}|n-l\rangle \tag{4.245}$$

Of course, this truncation is why the destruction operators were arranged to the right. The other operator may be taken to operate to the left and just produces the Hermitian conjugate of the above result:

$$\langle n|e^{u^*a^\dagger} = \sum_{m=0}^{n}\frac{(u^*)^m}{m!}\left[\frac{n!}{(n-m)!}\right]^{1/2}\langle n-m| \tag{4.246}$$

These two results must be multiplied together. Using the basic orthogonality of the states

$$\langle n-m|n-l\rangle = \delta_{n-m=n-l} = \delta_{m=l} \tag{4.247}$$

produces the compact result

$$\langle n|e^{u^*a^\dagger}e^{-ua}|n\rangle = \sum_{l=0}^{n}\frac{(-|u|^2)^l}{(l!)^2}\frac{n!}{(n-l)!} \tag{4.248}$$

This power series should be familiar to every student of physics: It is just the Laguerre polynomial of order n (remember hydrogen wave functions):

$$\langle n|e^{u^*a^\dagger}e^{-ua}|n\rangle = L_n(|u|^2) \tag{4.249}$$

The final step is to sum the series over n. This last series is just the generating function of Laguerre polynomials:

$$(1-z)\sum_{n=0}^{\infty}L_n(|u|^2)z^n = e^{|u|^2z/(z-1)} \tag{4.250}$$

In our case, to prove the theorem (4.240), identify

$$z = e^{-\beta\omega} \tag{4.251}$$

$$\frac{z}{z-1} = -N = -\frac{1}{e^{\beta\omega}-1} \tag{4.252}$$

When these factors are collected, the result does prove (4.240). The expression for $\mathscr{F}(t)$ is

$$\mathscr{F}(t) = e^{-\phi(t)} \tag{4.253}$$

$$\phi(t) = \lambda^2[(1-e^{-i\omega t}) + N|1-e^{i\omega t}|^2] \tag{4.254}$$

$$\phi(t) = \lambda^2[(N+1)(1-e^{-i\omega t}) + N(1-e^{i\omega t})] \tag{4.255}$$

Return to (4.222) and reintroduce the product over all $\mathbf{q}$ states. The function $F(t)$ contains a summation of the exponential factor:

$$F(t) = \Pi_{\mathbf{q}} \mathscr{F}_{\mathbf{q}}(t) = \exp\left[-\sum_{\mathbf{q}} \phi_{\mathbf{q}}(t)\right] \equiv \exp[-\Phi(t)] \tag{4.256}$$

$$\Phi(t) = \sum_{\mathbf{q}} \left(\frac{M_{\mathbf{q}}}{\omega_{\mathbf{q}}}\right)^2 [N_{\mathbf{q}}(1 - e^{i\omega_{\mathbf{q}}t}) + (N_{\mathbf{q}} + 1)(1 - e^{-i\omega_{\mathbf{q}}t})] \tag{4.257}$$

$$N_{\mathbf{q}} = \frac{1}{e^{\beta\omega_{\mathbf{q}}} - 1} \tag{4.258}$$

The final result for the particle Green's function for $t > 0$ is

$$G(t) = -ie^{-it(\varepsilon_c - \Delta)} e^{-\Phi(t)} (1 - n_F) \tag{4.259}$$

This equation is the exact result. Next proceed to a description of the physics.

4.3.3. Einstein Model

The physics is best understood by examining a simple application of the model. All the phonons are taken to have the same energy ω_0, which is called the Einstein model. The case of zero temperature will be discussed first. The modifications for nonzero temperature are derived afterwards. For zero temperature, all the phonon occupation factors are zero:

$$N_{\mathbf{q}} = N_0 = 0 \tag{4.260}$$

Furthermore, the summation over wave vector just produces a coupling constant g:

$$g = \sum_{\mathbf{q}} \frac{M_{\mathbf{q}}^2}{\omega_{\mathbf{q}}^2} \tag{4.261}$$

The Green's function still has the form (4.259), but now the factors are quite simple:

$$\Delta = g\omega_0 \tag{4.262}$$

$$\Phi(t) = g(1 - e^{-i\omega_0 t}) \tag{4.263}$$

The particle Green's function will be evaluated for the case of a single particle, so set $n_F = 0$. Evaluate

$$G(t) = -i\Theta(t) \exp[-it\varepsilon_c - g(1 - i\omega_0 t - e^{-i\omega_0 t})] \tag{4.264}$$

The same result is obtained for the retarded function. The spectral function is then the imaginary part of the retarded Green's function of frequency:

$$\begin{aligned} A(\omega) &= -2 \operatorname{Im}\left\{-i \int_0^\infty dt e^{i\omega t} \exp[-it\varepsilon_c - g(1 - i\omega_0 t - e^{-i\omega_0 t})]\right\} \\ &= 2 \operatorname{Re}\left\{\int_0^\infty dt e^{i\omega t} \exp[-it\varepsilon_c - g(1 - i\omega_0 t - e^{-i\omega_0 t})]\right\} \end{aligned} \tag{4.265}$$

The time integral may be evaluated by expanding the $ge^{-i\omega_0 t}$ part of the exponent in a power series,

$$\exp(ge^{-i\omega_0 t}) = \sum_l \frac{g^l}{l!} e^{-i\omega_0 tl} \tag{4.266}$$

so that the time integral contains terms such as

$$\int_0^\infty \exp[it(\omega - \varepsilon_c + \Delta - \omega_0 l)] = \frac{i}{\omega - \varepsilon_c + \Delta - \omega_0 l + i\delta} \tag{4.267}$$

The factor $i\delta$ is added to force the convergence of the oscillating integrand at large values of time. Then take the limit $\delta \to 0$ and obtain

$$\frac{i}{\omega - \varepsilon_c + \Delta - \omega_0 l + i\delta} = P\frac{i}{\omega - \varepsilon_c + \Delta - \omega_0 l} + \pi\delta(\omega - \varepsilon_c + \Delta - \omega_0 l)$$

The spectral function is the real part of this time integral, which has just the delta function:

$$A(\omega) = 2\pi e^{-g} \sum_{l=0}^{\infty} \frac{g^l}{l!} \delta(\omega - \varepsilon_c + \Delta - \omega_0 l) \tag{4.268}$$

The spectral function is a series of delta functions, spaced exactly ω_0 apart. The distribution of peak heights follows a Poisson distribution.

Remember that the spectral function is the probability that the particle has frequency ω. If there were no interactions, the particle would always have energy ε_c and there would be a single delta function at $\omega = \varepsilon_c$. This limit is obtained from (4.268) by setting $g = 0$. For $g \neq 0$ the particle has a nonzero probability of occupying other states which have l phonons. These configurations are not excited states. Since the temperature has been set equal to zero, the solution must be describing the ground state. In the ground state of the coupled system of particle and phonon, some probability exists that the system will have the different sets of frequencies $\omega_l \equiv \varepsilon_c - \Delta + l\omega_0$. The different values of ω_l obviously correspond to the particle being coupled to some phonons, which is certainly to be expected in this system. In the ground state, the particle energy fluctuates among these different values ω_l.

The spectral function (4.268) is shown in Fig. 4.7 for two different values of the coupling constant $g = 0.5$ and 5.5. These values correspond to the weak and strong coupling cases, respectively. Weak coupling is $g < 1$. Here the $l = 1$ peak is smaller than $l = 0$, and higher l peaks get smaller very rapidly. For strong coupling, the peak strength increases with l up to values of approximately $l \approx g$, and then the peaks decrease again.

It is useful to test the sum rules. The first one is [see (3.121)]

$$1 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A(\omega) = e^{-g} \sum_l \frac{g^l}{l!} \int_{-\infty}^{\infty} d\omega \delta(\omega - \omega_l) = e^{-g} \sum_l \frac{g^l}{l!} = 1$$

The factor $\exp(-g)$ in (4.268) is now recognized as the normalization factor which maintains the sum rule. These spectral functions also have the property that

$$\varepsilon_c = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \omega A(\omega) = \langle\omega\rangle \tag{4.269}$$

This moment is easily proved by direct evaluation,

$$\langle\omega\rangle = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \omega A(\omega) = e^{-g} \sum_l \frac{g^l}{l!} (\varepsilon_c - \Delta + \omega_0 l) = \varepsilon_c - \Delta + g\omega_0 = \varepsilon_c$$

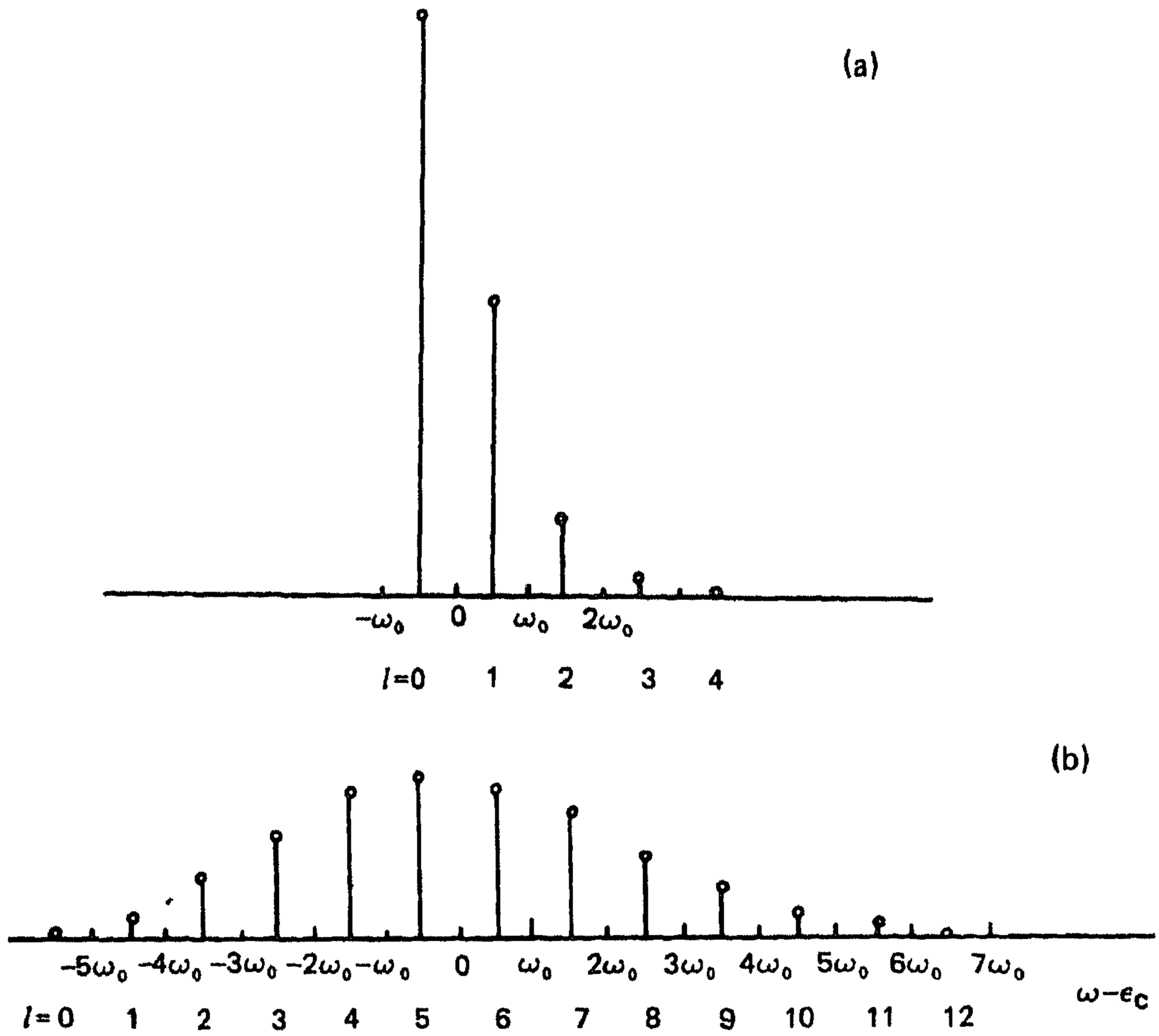


FIGURE 4.7 The spectral function of the independent boson model, Eqn. (4.270), shown for an Einstein model and two values of coupling constant. (a) For $g = 0.5$, (b) for $g = 5.5$.

which does give (4.269). The last integral is called the *first moment*. Higher moments $\langle\omega^n\rangle$ may also be evaluated, and they will depend on coupling constant. By using the relation

$$G_{\text{ret}}(t) = -i\Theta(t)\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}A(\omega)e^{-i\omega t} \tag{4.270}$$

the moments may also be shown to equal derivatives of the retarded Green's function:

$$\langle\omega^n\rangle = i\left[\frac{d^n}{dt^n}G_{\text{ret}}(t)\right]_{t\to 0^+} \tag{4.271}$$

Since the retarded Green's function is nonzero only at $t > 0$, one takes the time derivatives first and the limit of $t \to 0^+$ second.

An inspection of Fig. 4.7(b) verifies that the first moment is independent of coupling strength g. The delta function peaks have an intensity envelope which does appear to have a maximum near $\omega \approx \varepsilon_c$. As one increases the coupling constant g, the self-energy $\Delta = g\omega_0$ becomes larger in magnitude, so that the lowest energy peak shifts downward in energy. But its intensity lowers also because of the factor $\exp(-g)$, and the envelope of the delta function peaks becomes a Gaussian. A Poisson distribution becomes a Gaussian for large g, which

may be shown in the following way. Large g requires large values of l, so that Stirling's approximation may be used for the factorial:

$$l! \approx \sqrt{2\pi l}\, l^l e^{-l} \tag{4.272}$$

In this approximation the Poisson intensity becomes

$$\begin{aligned} e^{-g}\frac{g^l}{l!} &\approx \frac{1}{\sqrt{2\pi l}} \exp\left[-g + l + l\ln\left(\frac{g}{l}\right)\right] \\ &= \frac{1}{\sqrt{2\pi l}} \exp\left[-(g-l) + l\ln\left(1 + \frac{g-l}{l}\right)\right] \\ &\approx \frac{1}{\sqrt{2\pi l}} \exp\left[-(g-l) + l\left(\frac{g-l}{l}\right) - \frac{l}{2}\left(\frac{g-l}{l}\right)^2 + \cdots\right] \\ &\approx \frac{1}{\sqrt{2\pi l}} \exp\left[-\frac{(g-l)^2}{2l}\right] \end{aligned} \tag{4.273}$$

which is a Gaussian.

The zero-temperature result was described first, with an Einstein model, to emphasize that these additional peaks in the spectral function are really ground state properties of the system. For acoustical phonons, one gets a continuous distribution of possible phonon energies up to the bandwidth. The spectral function $A(\omega)$ becomes a continuous function of ω. The spectral shapes are dependent on whether the coupling to the acoustical phonons is piezoelectric or deformation potential.

Next consider the Einstein model at nonzero temperature. The phonon contribution $\Phi(t)$ in the Green's function is

$$\Phi(t) = g[(N+1)(1 - e^{-i\omega_0 t}) + N(1 - e^{i\omega_0 t})] \tag{4.274}$$

$$N = \frac{1}{e^{\beta\omega_0} - 1} \tag{4.275}$$

Because of the relationship

$$\frac{N+1}{N} = e^{\beta\omega_0}, \qquad \left(\frac{N+1}{N}\right)^{1/2} = e^{\beta\omega_0/2} \tag{4.276}$$

the exponent $\Phi(t)$ may have its terms grouped in the following fashion:

$$\Phi(t) = g\left\{2N + 1 - \sqrt{N(N+1)}(e^{-i\omega_0(t+i\beta/2)} + e^{i\omega_0(t+i\beta/2)}\right\} \tag{4.277}$$

The reason for this grouping will now become apparent. The objective is to expand $\exp[-\Phi(t)]$ in a power series in $\exp(i\omega_0 t)$, in order to evaluate the spectral function. Recall the series which generates the Bessel functions of complex argument, $I_n(z)$:

$$e^{z\cos(\theta)} = \sum_{l=-\infty}^{\infty} I_l(z) e^{il\theta} \tag{4.278}$$

The function $\exp[-\Phi(t)]$ has this form using

$$z = 2g\sqrt{N(N+1)} \tag{4.279}$$

$$\theta = \omega_0(t + i\beta/2) \tag{4.280}$$

The retarded Green's function at nonzero temperatures may be expanded as

$$G_{\text{ret}}(t) = -i\Theta(t)\ \exp[-it(\varepsilon_c - \Delta) - \Phi(t)] \tag{4.281}$$

$$= -i\Theta(t)\exp[-it(\varepsilon_c - \Delta) - g(2N+1)]$$

$$\times \sum_{l=-\infty}^{\infty} I_l\{2g\sqrt{N(N+1)}\}e^{l\omega_0(-it+\beta/2)} \tag{4.282}$$

and its spectral function is

$$A(\omega) = 2\pi\ \exp[-g(2N+1)] \sum_{l=-\infty}^{\infty} I_l\left\{2g\sqrt{N(N+1)}\right\}e^{l\omega_0\beta/2}$$

$$\times\ \delta(\omega - \varepsilon_c + \Delta - \omega_0 l) \tag{4.283}$$

The spectral function contains a summation over the frequencies $\omega_l = \varepsilon_c - \Delta + \omega_0 l$ and is a delta function at these values. In this sense it is similar to the zero-temperature result (4.268). Now the coefficient of the delta function is far more complicated. Another important difference is that negative values of l are permitted. Although $I_{-l} = I_l$, the factor $\exp[l\beta\omega_0/2)]$ skews the envelope of intensities to positive values of l.

4.3.4. Optical Absorption and Emission

The relaxation effects described above can, in some cases, actually be measured. Of course, one can never measure the properties of a one-particle Green's function, as was stated in Sec. 3.7. Linear response theory always gives a two-particle Green's function, which describes how the system responds when an external probe causes the system to change its state. However, in the many-boson model, there are some two-particle Green's functions which have properties nearly identical to the one-particle properties which have just been derived.

One important model is a localized defect with several possible localized electronic states. Each of these states may have a different matrix element for coupling to the phonon field:

$$H = \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} + \sum_i V_i \tag{4.284}$$

$$V_i = C_i^\dagger C_i \left[\varepsilon_i + \sum_{\mathbf{q}} M_{\mathbf{q}i}(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger)\right] \equiv h_i C_i^\dagger C_i \tag{4.285}$$

This Hamiltonian may also be exactly diagonalized. A canonical transformation of the previous form is done below:

$$\bar{H} = e^S H e^{-S}$$

$$S = \sum_i S_i = \sum_i C_i^\dagger C_i \sum_{\mathbf{q}} \frac{M_{\mathbf{q}i}}{\omega_{\mathbf{q}}}(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})$$

$$\bar{H} = \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} + \sum_i (\varepsilon_i - \Delta_i) C_i^\dagger C_i \tag{4.286}$$

$$\Delta_i = \sum_{\mathbf{q}} \frac{M_{\mathbf{q}i}^2}{\omega_{\mathbf{q}}} \tag{4.287}$$

Terms such as $n_i n_j$ will be set equal to zero, since it is assumed there is only one electron on the impurity. It may be in different states but not in two different electronic states at once. The Hamiltonian (4.284) was written with the electronic states not interacting with each other, except through the phonons. Any terms which permit a direct interaction between the states usually render the Hamiltonian unsolvable, at least exactly. For example, terms such as

$$[C_i^\dagger C_j + C_j^\dagger C_i] \sum_{\mathbf{q}} M_{\mathbf{q},ij}(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger) \tag{4.288}$$

are not included. They permit the particle to change its state by emitting a phonon. Such terms probably exist in real systems, but their addition to the Hamiltonian makes the problem much more difficult. They are customarily omitted.

In an optical absorption process, an electron may change its electronic state, say from i to j, by the absorption of a photon of frequency ω. This process is described by the Kubo formula, using the current–current correlation function. As in Sec. 4.3.3, the real-time correlation function may be employed. It is convenient to use the version of the Kubo formula given in Problem 16 of Chapter 3. For optical frequencies such that $\beta\omega \gg 1$, the formula is

$$\mathrm{Re}(\sigma_{\alpha\beta}) = \frac{1}{2\omega}\int_{-\infty}^{\infty} dt e^{i\omega t}\langle j_\alpha(t) j_\beta(0)\rangle \tag{4.289}$$

where (α, β) are (x, y, z) indices. The relation $\beta\omega \gg 1$ is easily satisfied, since typically $\hbar\omega_i \approx 2$–3 eV is in the visible spectrum, while at room temperature $\beta = 40\ \mathrm{eV}^{-1}$. For the transition between two localized states, the current operator is [see (1.194)]

$$j_\alpha = \sum_{ij} p_{ij,\alpha} C_i^\dagger C_j \tag{4.290}$$

The matrix element $p_{ij,\alpha}$ is treated as a constant in this problem. It plays no role in the many-body physics which follows. The correlation function is:

$$\begin{aligned} \langle j_\alpha(t) j_\beta(0)\rangle &= \sum_{ijkl} p_{ij,\alpha} p_{kl,\beta} \langle C_i^\dagger(t) C_j(t) C_k^\dagger C_l\rangle \\ U &= \langle C_i^\dagger(t) C_j(t) C_k^\dagger C_l\rangle = e^{\beta\Omega}\,\mathrm{Tr}(e^{-\beta H} e^{iHt} C_i^\dagger C_j e^{-iHt} C_k^\dagger C_l) \end{aligned} \tag{4.291}$$

The Hamiltonian H is that in (4.284). This correlation function may be solved exactly by using the same steps that were used to solve the Green's function (4.201) to get (4.259). First, the unit operator $1 = e^{-S}e^{S}$ is inserted into the trace:

$$U = e^{\beta\Omega}\,\mathrm{Tr}(e^{-\beta H} e^{iHt} C_i^\dagger C_j e^{-iHt} C_k^\dagger C_l e^{-S} e^{S}) \tag{4.292}$$

Using the cyclic property of the trace,

$$U = e^{\beta\Omega}\,\mathrm{Tr}(e^{S} e^{-\beta H} e^{iHt} C_i^\dagger C_j e^{-iHt} C_k^\dagger C_l e^{-S}) \tag{4.293}$$

The canonical transformation is taken on each term inside of the trace, which gives

$$U = e^{\beta\Omega}\,\mathrm{Tr}(e^{-\beta\bar{H}} e^{i\bar{H}t} X_i^\dagger X_j C_i^\dagger C_j e^{-i\bar{H}t} X_k^\dagger X_l C_k^\dagger C_l) \tag{4.294}$$

The factors X_i result from the transformation of the particle operators:

$$\bar{C}_i = e^S C_i e^{-S} = e^{S_i} C_i e^{-S_i} = C_i X_i \tag{4.295}$$

$$X_i = \exp\left[-\sum_{\mathbf{q}} \frac{M_{\mathbf{q}i}}{\omega_{\mathbf{q}}}(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})\right] \tag{4.296}$$

The transformed Hamiltonian H in (4.286) is diagonal in the operators C and $a_{\mathbf{q}}$. The time development of the correlation function may be found at once:

$$\begin{aligned} e^{i\bar{H}t} X_i^\dagger X_j C_i^\dagger C_j e^{-i\bar{H}t} &= X_i^\dagger(t) X_j(t) C_i^\dagger C_j e^{it(\varepsilon_i - \varepsilon_j - \Delta_i + \Delta_j)} \\ U = e^{it(\varepsilon_i - \varepsilon_j - \Delta_i + \Delta_j)} \mathrm{Tr}[e^{-\beta\bar{H}} X_i^\dagger(t) X_j(t) X_k^\dagger(0) X_l(0) C_i^\dagger C_j C_k^\dagger C_l] \end{aligned} \tag{4.297}$$

The electron and phonon parts of the trace may be separated, which is permissible because the X_i operators do not depend on particle states:

$$U = e^{it(\varepsilon_i - \varepsilon_j - \Delta_i + \Delta_j)} U_{el} U_{ph}(t) \tag{4.298}$$

$$U_{el} = \mathrm{Tr}[e^{-\beta\bar{H}} C_i^\dagger C_j C_k^\dagger C_l] \tag{4.299}$$

$$U_{ph}(t) = \mathrm{Tr}[e^{-\beta\bar{H}} X_i^\dagger(t) X_j(t) X_k^\dagger(0) X_l(0)] \tag{4.300}$$

The electron part is quickly evaluated. Using Wick's theorem, the subscripts referring to particle states must be paired. In fact, one must have that $j = k$ and $i = l$:

$$U_{el} = \mathrm{Tr}[e^{-\beta\bar{H}} C_l^\dagger C_k C_k^\dagger C_l] = n_l(1 - n_k) \tag{4.301}$$

$$U_{ph}(t) = \mathrm{Tr}[e^{-\beta\bar{H}} X_l^\dagger(t) X_k(t) X_k^\dagger(0) X_l(0)] \tag{4.302}$$

This result is useful when evaluating the phonon part of the trace. It is similar in form to the earlier trace for the one-particle Green's function. The only difference is that there are now four factors instead of two. The evaluation procedure is the same, since the four operators can be paired and combined into two. This simplification happens because operators which are at the same time can be combined, since their exponents commute:

$$X_l^\dagger(t) X_k(t) = e^{iHt}\left(\Pi_{\mathbf{q}} e^{\lambda_l(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})} e^{-\lambda_k(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})}\right) e^{-iHt} \tag{4.303}$$

$$= e^{iHt}\left(\Pi_{\mathbf{q}} e^{(\lambda_l - \lambda_k)(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})}\right) e^{-iHt} \tag{4.304}$$

$$= \exp\left[\sum_{\mathbf{q}} \frac{M_{\mathbf{q}l} - M_{\mathbf{q}k}}{\omega_{\mathbf{q}}}(a_{\mathbf{q}}^\dagger e^{t\omega_{\mathbf{q}}t} - a_{\mathbf{q}} e^{-i\omega_{\mathbf{q}}t})\right] \tag{4.305}$$

It is only possible to add two exponential operators in this fashion when they commute. They do in this case when taken at the same time. The correlation function $U_{ph}(t)$ above is immediately simplified to

$$U_{ph}(t) = \Pi_{\mathbf{q}}\, \mathrm{Tr}\{e^{-\beta H} \exp[-\lambda(a_{\mathbf{q}}^\dagger e^{i\omega_{\mathbf{q}}t} - a_{\mathbf{q}} e^{i\omega_{\mathbf{q}}t})] e^{\lambda(a_{\mathbf{q}}^\dagger - a_{\mathbf{q}})}\} \tag{4.306}$$

$$\lambda = -\frac{M_{\mathbf{q}l} - M_{\mathbf{q}k}}{\omega_{\mathbf{q}}} \tag{4.307}$$

This equation for $U_{ph}(t)$ is precisely the form (4.222) which was untangled previously, and now the effective coupling constant is $\lambda = [M_{\mathbf{q}k} - M_{\mathbf{q}l}]/\omega_{\mathbf{q}}$.

The previous untangling result (4.257) can be used to obtain the evaluation of $U_{ph}(t)$:

$$U_{ph}(t) = \exp\{-\Phi_{kl}(t)\}$$

$$\Phi_{kl}(t) = \sum_{\mathbf{q}} \frac{(M_{\mathbf{q}l} - M_{\mathbf{q}k})^2}{\omega_{\mathbf{q}}^2}[(N_{\mathbf{q}} + 1)(1 - e^{-i\omega_{\mathbf{q}}t})] + N_{\mathbf{q}}(1 - e^{i\omega_{\mathbf{q}}t})]$$

These various results are collected, and the result for the correlation function for the conductivity is

$$\mathrm{Re}[\sigma_{\alpha\beta}(\omega)] = \frac{1}{2\omega}\sum_{kl} n_l(1 - n_k)p_{kl,\alpha}p^*_{kl,\beta}$$
$$\times \int_{-\infty}^{\infty} dt\ \exp\{it[\omega + \varepsilon_l - \varepsilon_k - \Delta_l + \Delta_k - \Phi_{kl}(t)]\} \tag{4.308}$$

The function Φ_{kl} in the exponent is similar to the one for the particle Green's function. The only difference is that the effective matrix element is the difference between the two single-state matrix elements: $M_{kl,\mathbf{q}} = M_{\mathbf{q}k} - M_{\mathbf{q}l}$. For example, if the two matrix elements $M_{\mathbf{q}k}$ and $M_{\mathbf{q}l}$ happen to be equal, then the effective matrix element vanishes. In this unlikely circumstance the spectral function is just a delta function at the frequency $\omega = \varepsilon_k - \varepsilon_l$. Usually $M_{\mathbf{q}k}$ and $M_{\mathbf{q}l}$ are not equal, at least not for all different wave vectors, so that phonon effects are present in the transition.

The above model describes dynamic relaxation. In the initial state of the system, the electron is in a state l, and the phonons are relaxed about their equilibrium configuration for the state l. Recall that if the phonon part of the Hamiltonian is written in terms of harmonic oscillator coordinates, the phonons relax to an equilibrium configuration given by [see (1.245)]

$$Q_{\mathbf{q}}^{(l)} = -2\left(\frac{\hbar}{2\rho\omega_{\mathbf{q}}}\right)^{1/2}\frac{M_{\mathbf{q}l}}{\omega_{\mathbf{q}}} \tag{4.309}$$

In the optical absorption, the electron starts in state l and ends in state k. The phonons start with an equilibrium configuration about the point $Q_{\mathbf{q}}^{(l)}$ but end the optical transition with the equilibrium configuration about the point $Q_{\mathbf{q}}^{(k)}$. The phonon system must alter its equilibrium configuration during the transition. This change is a relaxation process, since it must relax to the new equilibrium configuration during the optical step. The physics problem is to determine how the phonon relaxation process affects the absorption spectra. This information is contained in the result (4.308).

The process is indicated schematically in Fig. 4.8, which shows a potential energy diagram for each oscillator coordinate $Q_{\mathbf{q}}$. There are two parabolic curves, with parabolicity $\omega_{\mathbf{q}}^2$. The lower curve describes the ground state of the system. The electronic energy is ε_l. If there were no coupling to the particle in state l, the phonon parabola would be a minimum at this point $Q_{\mathbf{q}} = 0$. However, because of the coupling $M_{\mathbf{q}l}$ to this particle, the potential minimum is at $Q_{\mathbf{q}}^{(l)}$. The upper curve describes the final state potential energy curves of the phonons plus particle. The particle energy is ε_k, and the curve crosses here because the phonon system has this energy when $Q_{\mathbf{q}} = 0$. This potential energy curve has the minimum at $Q_{\mathbf{q}}^{(k)}$. This minimum has been drawn on the other side of the axis to emphasize that it is usually a point different than the ground state minimum. Figure 4.8 is called a *configurational coordinate diagram* (Williams and Hebb, 1951). Such diagrams were originally constructed with a single-particle coordinate Q, which represented the real atom displacement of the first

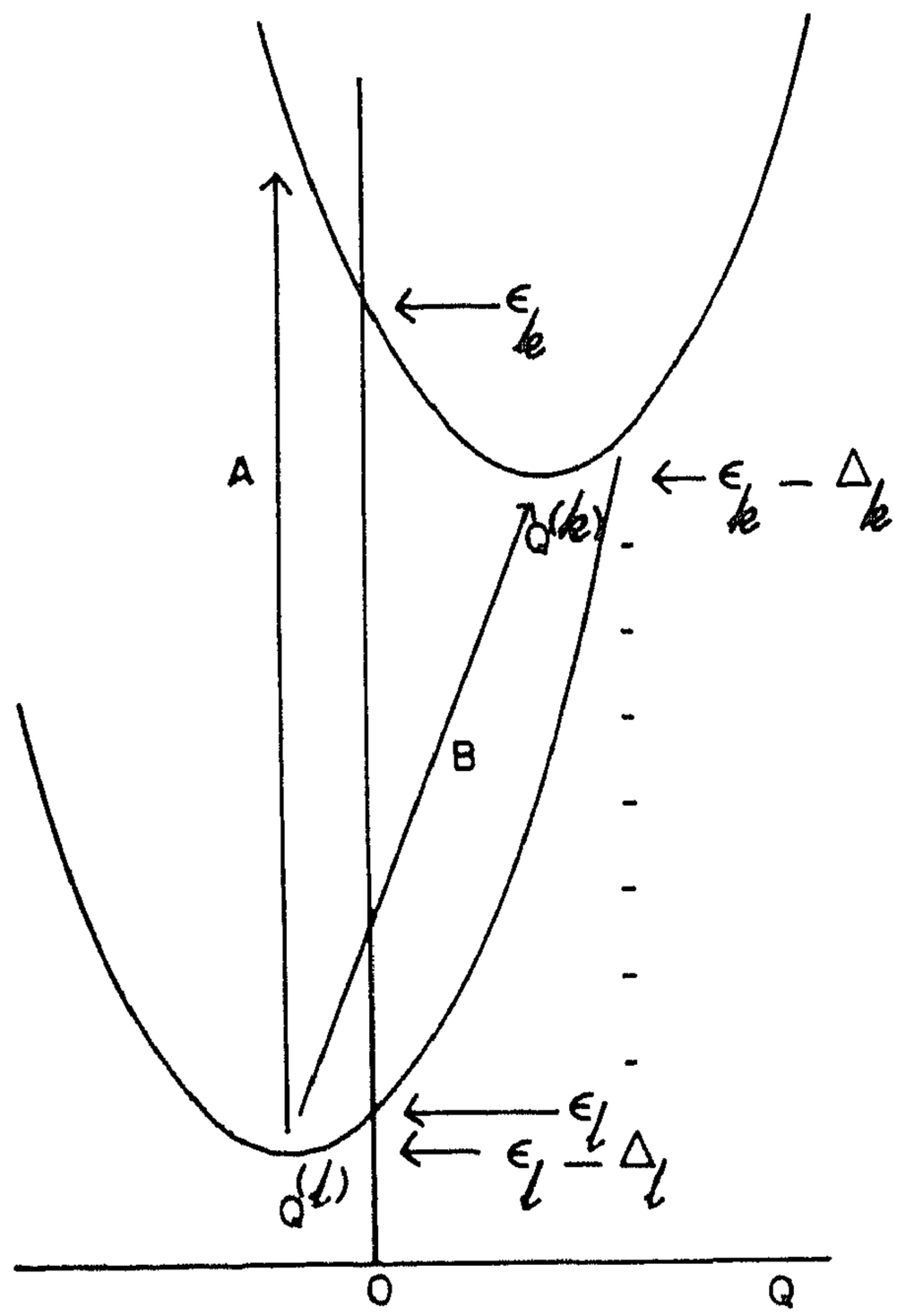


FIGURE 4.8 Configurational coordinate drawing of the independent boson model. The two parabolas represent the phonon potential energy of the initial and final electronic states in the transition. Transition path A is most likely, while path B is less likely but gives the zero-phonon probability.

shell of atoms about the impurity. However, the use of configurational coordinate diagrams is much more rigorous when interpreted as the potential energy of the individual phonon modes.

The physical question is whether the optical transition happens vertically on this diagram or along some other trajectory. For example, the arrow B is from one potential minimum to the other. This transition is called the *zero-phonon line.* In reality, all these transitions are possible, and each has a probability of occurrence. These probabilities are given by (4.308). The physics is well illustrated by again using the Einstein model for phonons. This formula is expanded in the same manner used to derive the one-particle Green's function (4.283):

$$\mathrm{Re}\ \sigma_{\alpha\beta}(\omega) = \frac{\pi}{\omega}\sum_{kl} n_l(1-n_k)p_{kl,\alpha}p^*_{kl,\beta}e^{-g_{kl}(T)} \sum_{m=-\infty}^{\infty} \delta(\omega-\omega_m)I_m(\gamma_{kl})e^{m\beta\omega_0/2}$$

$$g_{kl}(T) = (2N+1)g_{kl} = (2N+1)\sum_{\mathbf{q}} \frac{(M_{\mathbf{q}k}-M_{\mathbf{q}l})^2}{\omega_{\mathbf{q}}^2} \tag{4.310}$$

$$\omega_m = \varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l + m\omega_0 \tag{4.311}$$

$$\gamma_{kl} = 2g_{kl}\sqrt{N(N+1)} \tag{4.312}$$

This formula for $\mathrm{Re}[\sigma(\omega)]$ has exactly the same form as the one-particle spectral function (4.283), as an expansion in Bessel functions of complex argument. Now the coupling constant to the phonons is determined by $M_{\mathbf{q}k} - M_{\mathbf{q}l}$. If $M_{\mathbf{q}k} = M_{\mathbf{q}l}$, then $g = 0, \gamma = 0$, and the conductivity is a single delta function at $\omega = \varepsilon_k - \varepsilon_l$.

The zero-phonon line is given by the term with $m = 0$. It has the energy $\omega_{m=0} = \varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l$, which corresponds to the B arrow in Fig. 4.8. The probability of this transition is proportional to

$$e^{-g_{kl}(T)} I_0(\gamma_{kl}) \tag{4.313}$$

The envelope of delta function heights is determined by the m dependence of the factor $I_m(\gamma_{kl}) \exp(m\beta\omega_0/2)$. For coupling constants g_{kl} greater than unity, the delta function intensifies with m and then falls off at higher values. The strongest delta function occurs at positive nonzero values of m, which is illustrated in Fig. 4.9. The circles show the peak heights calculated for $g_{kl} = 4$ and $N = \frac{1}{2}$. The points are the emission spectra, which will be derived below.

The usual interpretation is that terms with $m > 0$ correspond to the creation of m phonons during the optical transition. The zero-phonon line is just the transition where no phonons are created or destroyed. Transitions in which $m > 0$ are phonon emission, and transitions in which $m < 0$ are phonon absorption. The phonon absorption process is possible at nonzero temperature, when phonons are thermally present in the initial system. This interpretation ignores the fact that both the ground and excited states are an admixture of phonon states. The spectral functions (4.283) of both initial and final states contain admixtures of different numbers of phonons. These differing admixtures are due to fluctuations in the system. Sometimes the particle has one phonon around it, while other times it has three or four, etc. The probability of different fluctuations is given by the Bessel function coefficients in (4.283).

The other interpretation is that the absorption spectra of Fig. 4.9 correspond to transitions between states of different fluctuations. The initial state may be in a state with a fluctuation of plus four phonons and the final state with plus six. It takes a net of plus two phonon energies to complete the transition with energy conservation. This model is not really different from the above interpretation. The phonons in the fluctuation still have to dissipate away. However, one state may have a different number of average phonons, in its fluctuating

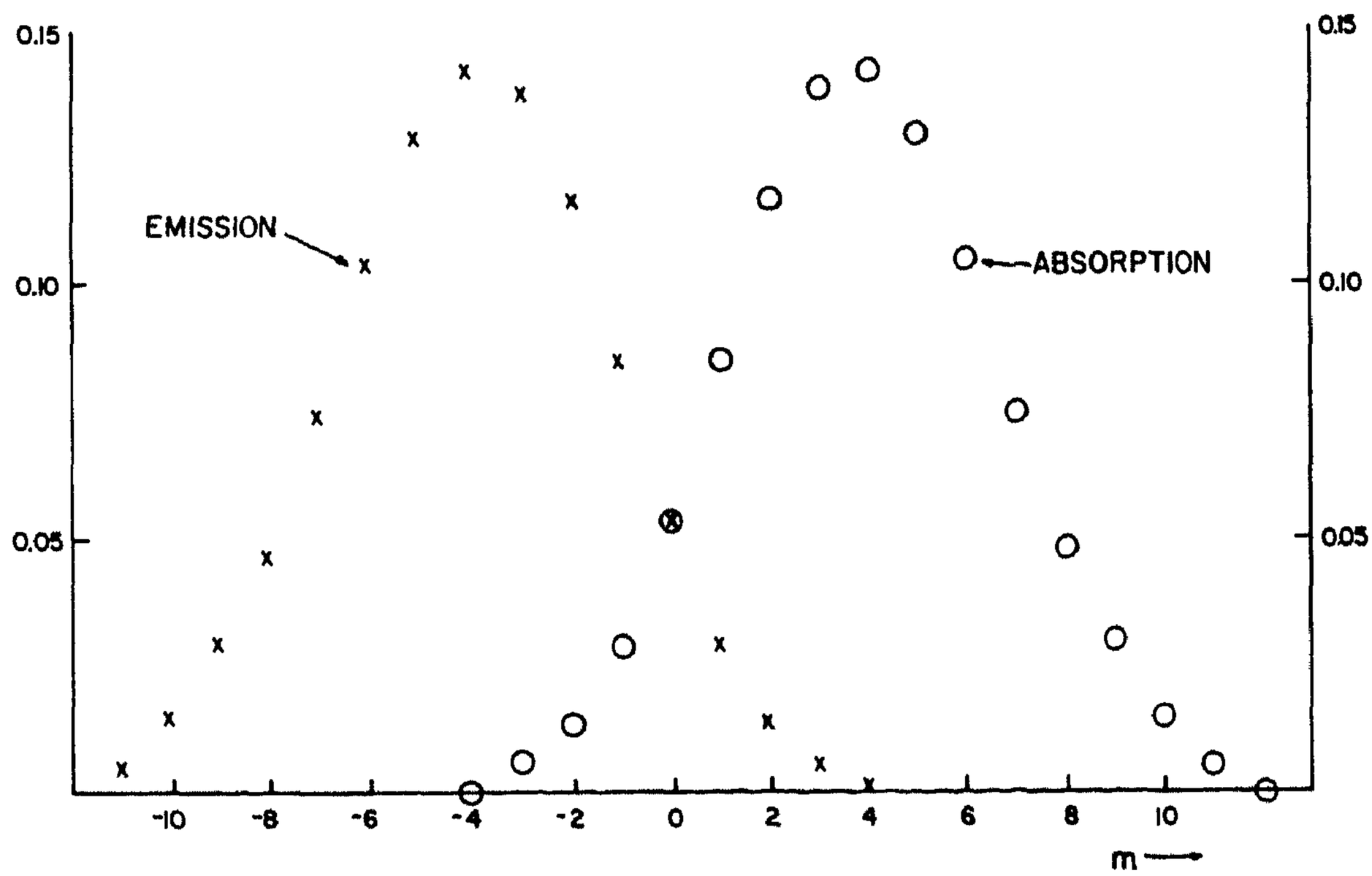


FIGURE 4.9 The emission and absorption spectra for an Einstein model, in the independent boson approximation, from Eq. (4.239). The o's and ×'s mark the peak intensities of each phonon sideband.

cloud, than the other. The number of phonons which must be created or destroyed to make up this average is not really creating a net amount to dissipate into the system. They will, instead, stay at the impurity site and take part in the fluctuations.

The emission spectra will now be considered. It is assumed that the particle has been elevated to an excited state at an earlier time, perhaps by an optical absorption process. However, experimentally it is often done by electron bombardment. The phonons in the excited state relax around the particle and eventually attain the equilibrium configuration plus fluctuations. This assumption is valid only for long-lived excited states. Then a photon is emitted while the particle drops to a lower energy state. Of course, the phonon system must adjust during the emission process just as it did during the absorption. The same relaxation processes are encountered again.

The emission spectra is derived without doing any more work. It may be obtained from the absorption result by a simple argument. The emission is the rate per unit time that photons are produced in the solid. Of course, the absorption is the rate at which they disappear. Write a simple rate equation for the average number of photons N of frequency ω,

$$\frac{dN}{dt} = -wN \tag{4.314}$$

which has the solution

$$N(t) = N_i e^{-wt} \tag{4.315}$$

where w is the rate of absorption. Since the conductivity $\sigma(\omega)$ also has the units of s^{-1}, one might expect that $w \propto \sigma$. In fact, the exact relationship is

$$w = \frac{4\pi\sigma}{\varepsilon_1} \tag{4.316}$$

where ε_1 is the square of the refractive index. The rate of photon absorption is just proportional to the real part of the conductivity, as given by the Kubo formula. Another way to understand this is to consider rewriting the Kubo formula with some additional factors:

$$\sigma(\omega) \to (1 - e^{-\beta\omega}) \int_{-\infty}^{\infty} dt e^{i\omega t} \langle j(t) b^\dagger(t) j(0) b(0) \rangle \tag{4.317}$$

The two additional factors are the creation and destruction operators for photons b and $b^\dagger$. The right-hand part of the correlation factor,

$$j_\alpha(0) b(0) = \sum_{kl} p_{kl,\alpha} C_k^\dagger C_l b \tag{4.318}$$

describes the process whereby a photon is destroyed (by being absorbed) and an electron has its state changed from l to k. This term arises from the $p \cdot A$ interaction. The left-hand part of the correlation function contains the factors

$$j_\beta(t) b^\dagger(t) = \sum_{kl} p_{kl,\beta} C_l^\dagger(t) C_k(t) b^\dagger(t) \tag{4.319}$$

which describe the inverse process at another time t. The correlation function between these two events is the Kubo formula. The photon parts may be factored out of the expression using $b^\dagger(t) = b^\dagger e^{i\omega t}$ and $\langle b^\dagger b \rangle = N$

$$\sigma_{\beta\alpha}(\omega) \propto N(1 - e^{-\beta\omega}) \int_{-\infty}^{\infty} dt e^{i\omega t} \langle j_\beta(t) j_\alpha(0) \rangle \tag{4.320}$$

This quantity is, except for the factor of 4π, just the right-hand side of Eq. (4.314). It is the rate that photons are destroyed in the sample by optical absorption.

For emission, calculate the rate at which photons are created. The first step in the correlation function would describe a photon being created while the electron changes its state. This correlation function should be

$$\int_{-\infty}^{\infty} dt \langle j_\alpha(t) b(t) j_\beta(0) b^\dagger(0) \rangle \tag{4.321}$$

where the prefactor $(1 - e^{-\beta\omega})$ is dropped since it is near unity. The first operator on the right $b^\dagger$ creates a photon as part of the emission process. Take these photon factors out of the integrand, using $b(t) = be^{-i\omega t}$ and $\langle bb^\dagger \rangle = N + 1$, to obtain

$$\langle bb^\dagger \rangle \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle j_\alpha(t) j_\beta(0) \rangle = (N+1) \frac{I(\omega)}{4\pi} \tag{4.322}$$

This result, multiplied by 4π, must be the rate that photons are being made in the solid:

$$\frac{dN}{dt} = I(N+1) \tag{4.323}$$

$$I(\omega) = 4\pi \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle j_\beta(t) j_\alpha(0) \rangle \tag{4.324}$$

The formula for $I(\omega)$ is identical with the Kubo formula except that the frequency has changed sign. Compare this formula with the Einstein model result (4.310), for example, and only change the sign of ω to get

$$I(\omega) \propto e^{-g(2N+1)} \sum_{m=-\infty}^{\infty} \delta(\omega + \omega_m) I_m(\gamma) e^{m\beta\omega_0/2} \tag{4.325}$$

The outcome is very simple. The emission spectra is just the mirror image of the absorption spectra reflected across the zero-phonon line. The emission is illustrated by the points marked X in Fig. 4.9. The emission spectra are mostly on the low-frequency side of the zero-phonon line. In general, the emission spectra have a lower average frequency than the corresponding absorption spectra for the same processes. The difference is a consequence of relaxation, which is illustrated using the configurational coordinate diagram of Fig. 4.10. The vertical arrow "abs" shows the most likely absorption event. Of course, other absorption events are possible, but the peak of the envelope of delta functions is at this vertical transition.

Similarly, the most likely emission event is the downward arrow marked "emiss." It is shorter than the absorption arrow, which indicates that the average emission takes less energy. No matter how one draws those two parabolas with respect to each other, the downward emission arrow is always shorter than (or equal to) the upward arrow. This fact can also be demonstrated by using the various theoretical formulas which have been derived. However, the simple diagram of Fig. 4.10 is probably the clearest proof that the emission has an average lower energy than absorption.

The lower average energy in emission, compared to absorption, is understood using the first interpretation mentioned after (4.310). During absorption, one usually makes phonons. The photon energy is given by the change in particle energy plus the phonon energy

$$\omega = \Delta(\text{particle energy}) + \text{phonons} \tag{4.326}$$

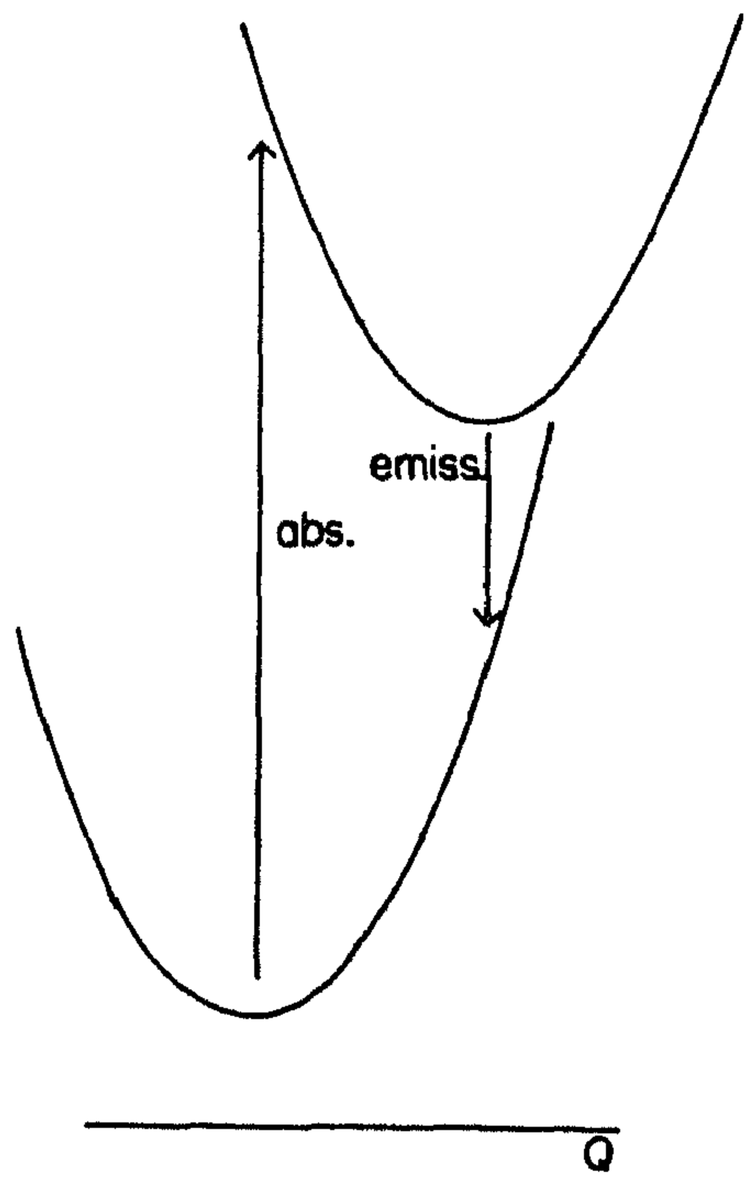


FIGURE 4.10 Configurational coordinate drawing shows that the average absorption frequency is higher than the average emission frequency.

This energy is higher than the particle transition energy. In the emission event, the energy comes from the particle transition, but here too one makes phonons. Energy conservation is

$$\Delta(\text{particle energy}) = \omega + \text{phonons} \tag{4.327}$$

and the average photon energy ω is less than the particle transition energy.

Figure 4.11 shows an experimental result for absorption and emission at an impurity center in ZnTe, reported by Dietz *et al.* (1962). The emission and absorption are mirror images about the zero-phonon lines. The arrows at the top show the separation of the optical phonon lines, which are very clear in the emission spectra. The two other peaks between successive LO phonons are sidebands due to LA and TA phonons. These data were taken at 20 K. It is an unusually good example of the relation between emission and absorption.

4.3.5. Sudden Switching

There is another way to derive the result for optical emission and absorption. Of course it obtains the same answer, since both derivations are exact. This other derivation emphasizes the switching aspects of the problem. It is just an alternate method of obtaining the same answer but perhaps provides a slightly different physical insight. The many-body problem is presented as the system response when a potential is suddenly switched on. Return to the Kubo formula and consider the time correlation function:

$$U(t) = \mathrm{Tr}(e^{-\beta H} e^{iHt} C_l^\dagger C_k e^{-iHt} C_k^\dagger C_l) \tag{4.328}$$

In writing this expression, the subscripts of the particle operators are paired up. Now the Hamiltonian H occurs at three places in this expression. It has two different forms: one for each different place it occurs in the sequence of operators. The factor on the right is

$$e^{-iHt} C_k^\dagger C_l \tag{4.329}$$

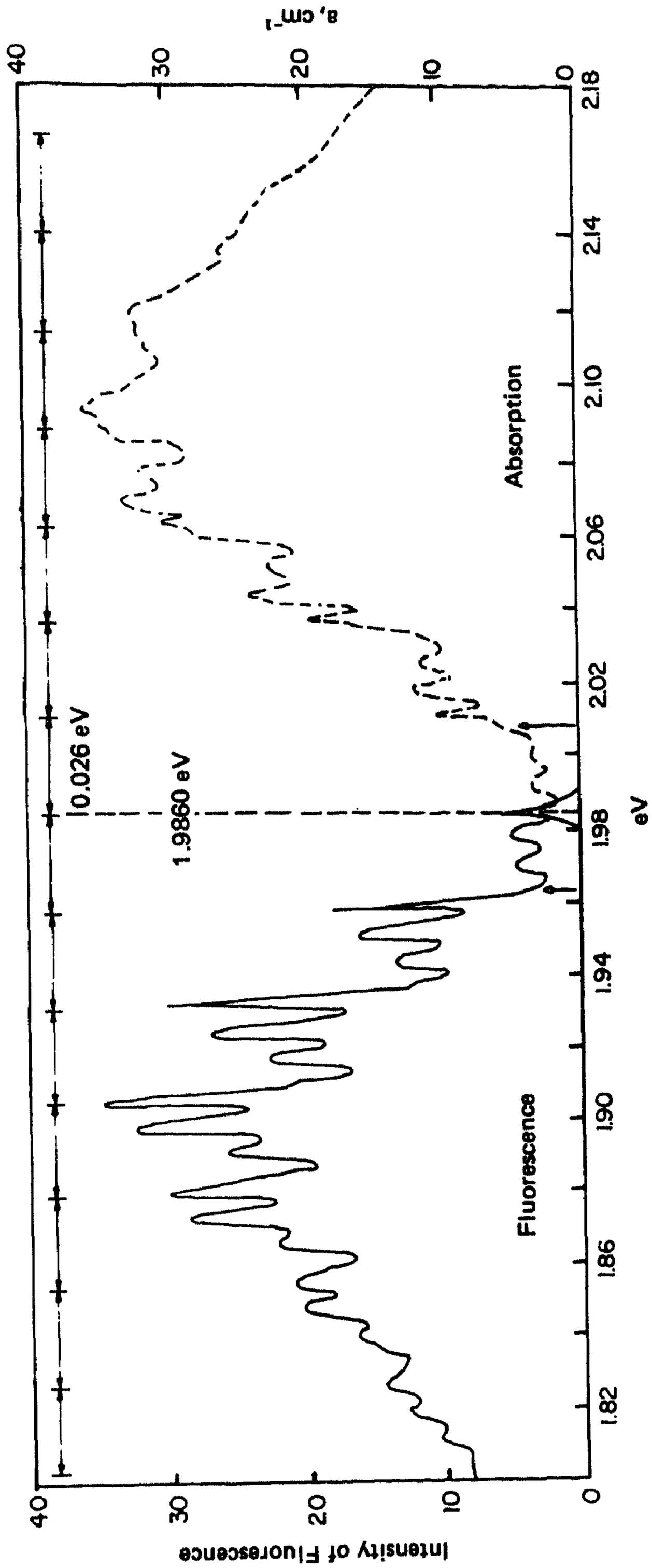


FIGURE 4.11 The fluorescence and absorption at 20 K from an impurity center in ZnTe. The positions of the peak are mirrored about a central no-phonon line common to both spectra. The LO phonon energy intervals are indicated. The other two peaks in each interval are from TA and LA phonons. *Source:* Dietz *et al.* (1962) (used with permission).

The operators $C_k^\dagger C_l$ act to the right. They destroy the particle in state l and create one in state k. The operator H acts on a system with a particle in k but not in l. According to the form (4.284) for the Hamiltonian,

$$H = H_p + \sum_i h_i C_i^\dagger C_i \tag{4.330}$$

$$H_p = \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \tag{4.331}$$

When H operates on a state $|k\rangle = C_k^\dagger|0\rangle$ then $H|k\rangle = (H_p + h_k)|k\rangle$. In this case, $e^{-iHt}C_k^\dagger C_l$ may be replaced in the operator sequence by

$$e^{-iHt}C_k^\dagger C_l = e^{-it(H_p+h_k)}C_k^\dagger C_l \tag{4.332}$$

This result is exact, and no approximation is involved in the replacement. The other two factors of H are at the left of the operator sequence. On their right is the immediate sequence of particle operators,

$$e^{-\beta H}e^{iHt}C_l^\dagger C_k \tag{4.333}$$

which returns the particle back to the state l. These two Hamiltonians operate on the system with a particle in the state l and therefore produce

$$e^{-\beta(H_p+h_l)}e^{it(H_p+h_l)}C_l^\dagger C_k \tag{4.334}$$

The correlation function $U(t)$ in (4.328) is exactly equal to

$$U(t) = \mathrm{Tr}(e^{-\beta(H_p+h_l)}e^{it(H_p+h_l)}C_l^\dagger C_k e^{-it(H_p+h_k)}C_k^\dagger C_l) \tag{4.335}$$

The particle operators may now be removed from the trace. The time development of the correlation function involves only the operators H_p and h_j, and these contain only phonon operators. The particle operators have done their job, in the correlation function, by determining which effective Hamiltonians operate at various points in the sequence. There remains only to evaluate the phonon part of the trace:

$$U(t) = e^{it(\varepsilon_l-\varepsilon_k)}n_l(1-n_k)U_{ph}(t) \tag{4.336}$$

$$U_{ph}(t) = \mathrm{Tr}(e^{-\beta(H_p+\tilde{h}_l)}e^{it(H_p+\tilde{h}_l)}e^{-it(H_p+\tilde{h}_k)}) \tag{4.337}$$

$$\tilde{h}_j = \sum_{\mathbf{q}} M_{\mathbf{q}j}(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger) \tag{4.338}$$

This correlation function could be evaluated by Feynman disentangling. However, another method will now be developed.

When a Hamiltonian is solved in the interaction representation one writes $H = H_0 + V$, where H_0 is a part one can solve and V is the perturbation. Here the procedure is slightly different. The Hamiltonian H_0 is for the state with a particle in l,

$$H_0 = H_p + h_l \tag{4.339}$$

and $H_0 + V$ is the Hamiltonian for the particle in the state k,

$$H_0 + V = H_p + h_k \tag{4.340}$$

$$V = h_k - h_l = \varepsilon_k - \varepsilon_l + \sum_{\mathbf{q}}(M_{\mathbf{q}k} - M_{\mathbf{q}l})(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger) \tag{4.341}$$

The perturbation V is the difference between these two Hamiltonians. The phonon part of the correlation function may be written as

$$U_{ph}(t) = \mathrm{Tr}(e^{-\beta H_0} e^{itH_0} e^{-it(H_0+V)}) \tag{4.342}$$

The next step is to make H_0 a diagonal Hamiltonian, which is not difficult, since the unitary transformation is given earlier in (4.286):

$$\begin{aligned}
\bar{H}_0 &= e^{S_l} H_0 e^{-S_l} = \varepsilon_l - \Delta_l + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \\
\bar{H}_0 + \bar{V} &= e^{S_l}(H_0 + V)e^{-S_l} = \varepsilon_k - \Delta_k + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \\
&\quad + \sum_{\mathbf{q}} \frac{(M_{\mathbf{q}k} - M_{\mathbf{q}l})^2}{\omega_{\mathbf{q}}} + \sum_{\mathbf{q}} (M_{\mathbf{q}k} - M_{\mathbf{q}l})(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger) \\
S_l &= \sum_{\mathbf{q}} \frac{M_{\mathbf{q}l}}{\omega_{\mathbf{q}}} (a_{\mathbf{q}}^\dagger - a_{\mathbf{q}}) \\
\Delta_{kl} &= \sum_{\mathbf{q}} \frac{(M_{\mathbf{q}k} - M_{\mathbf{q}l})^2}{\omega_{\mathbf{q}}}
\end{aligned} \tag{4.343}$$

The transformed potential $\bar{V}$ is obtained by subtracting $\bar{H}_0$ from $\bar{H}_0 + \bar{V}$:

$$\bar{V} = \varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l + \Delta_{kl} + \delta\bar{V} \tag{4.344}$$

$$\delta\bar{V} = \sum_{\mathbf{q}} (M_{\mathbf{q}k} - M_{\mathbf{q}l})(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger) \tag{4.345}$$

The unitary transformation in (4.343) is applied to all the operators in the correlation function (4.342). It is done following the same steps used to transform (4.291), except that here only $\exp(S_l)$ is used,

$$U_{ph}(t) = \mathrm{Tr}(e^{-\beta \bar{H}_0} e^{it\bar{H}_0} e^{-it(\bar{H}_0+\bar{V})}) \tag{4.346}$$

where $\bar{H}_0$ now has the diagonal form in (4.343).

This correlation function has the form of the switching phenomena, as mentioned at the beginning of this subsection. It describes the response of the system to suddenly switching on the potential $\bar{V}$ at time $t = 0$. This switching on causes numerous transients in the phonon system, and these transients are the cause of the phonon sidebands observed in absorption.

To illustrate this switching, consider a system which has a Hamiltonian given by

$$H = \bar{H}_0 + \bar{V} d^\dagger d \tag{4.347}$$

where $d^\dagger$ and d are creation and destruction operators for some particle. The Green's function of the d particle is given for $(t > 0)$

$$G_d(t) = -i\langle 0|e^{iHt} d e^{-iHt} d^\dagger|0\rangle \tag{4.348}$$

The particle d is created at time $t = 0$. Then e^{-iHt} operates on the state with a particle in d and $Hd^\dagger|0\rangle = (\bar{H}_0 + \bar{V})d^\dagger|0\rangle$. At a later time t the particle is destroyed by the destruction operator d, and e^{iHt} operates on a state with no d-particle. The Green's function for $t > 0$ is

$$G_d(t) = -i\langle dd^\dagger\rangle \langle e^{i\bar{H}_0 t} e^{-it(\bar{H}_0+\bar{V})}\rangle \tag{4.349}$$

which is exactly the form of the correlation function (4.346). All the d particle did for us was to switch on the potential $\bar{V}$ at $t = 0$ and switch it off at time t. This switching-on causes the phonons to respond and adjust to the new potential $\bar{V}$. These transients are the same relaxation processes which were discussed earlier. The correlation function (4.346) is evaluated by first recalling the interaction representation result (2.1.7):

$$e^{i\bar{H}_0 t} e^{-it(\bar{H}_0+\bar{V})} = T \exp\left[-i \int_0^t dt_1 \bar{V}(t_1)\right] \tag{4.350}$$

In this case, the operator $\bar{V}(t)$ is

$$\bar{V}(t) = e^{i\bar{H}_0 t} \bar{V} e^{-i\bar{H}_0 t} \tag{4.351}$$

$$= \varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l + \Delta_{kl} + \delta\bar{V}(t) \tag{4.352}$$

$$\delta\bar{V}(t) = \sum_{\mathbf{q}} (M_{\mathbf{q}k} - M_{\mathbf{q}l})(a_{\mathbf{q}} e^{-i\omega_{\mathbf{q}} t} + a_{\mathbf{q}}^\dagger e^{i\omega_{\mathbf{q}} t}) \tag{4.353}$$

The constant terms may be immediately removed from the correlation function. Only the operator part $\delta\bar{V}(t)$ needs to be time-ordered:

$$e^{i\bar{H}_0 t} e^{-it(\bar{H}_0+\bar{V})} = \exp[-it(\varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l + \Delta_{kl})] T \exp\left[-i \int_0^t dt_1 \delta\bar{V}(t_1)\right]$$

Now $\delta\bar{V}(t_1)$ contains only two terms, which are, respectively, proportional to $a^\dagger$ and a. Since the commutator of these results is a constant, one immediately considers evaluating these time-ordered exponents by using Feynman's theorem (4.226). It is not correct to use precisely the form of (4.226). The two terms in the exponent operate at different times, and this fact must be considered. The proper procedure is done in Problem 5 of Chapter 2, which is to separate the two terms in the time-ordered exponent. Again it is done for a single value of wave vector $\mathbf{q}$, and this subscript is suppressed. The equations are:

$$T \exp\left[-i \int_0^t dt_1 \delta\bar{V}(t_1)\right] = e^{a^\dagger \phi(t)} \exp\left[-i\lambda \int_0^t dt_1 e^{-a^\dagger \phi(t_1)} a e^{a^\dagger \phi(t_1)} e^{-i\omega t_1}\right]$$

$$\phi(t) = -i\lambda \int_0^t dt_1 e^{i\omega t_1} = \frac{\lambda}{\omega}(1 - e^{i\omega t}) \tag{4.354}$$

$$\lambda = (M_l - M_k)$$

The time dependence of the exponent in the integrand is

$$e^{-a^\dagger \phi(t_1)} a e^{a^\dagger \phi(t_1)} = a + \phi(t_1) \tag{4.355}$$

This time integral may now be evaluated and gives

$$-i\lambda \int_0^t dt_1 e^{-i\omega t}\left[a + \frac{\lambda}{\omega}(1 - e^{i\omega t})\right] \tag{4.356}$$

$$= -a\frac{\lambda}{\omega}(1 - e^{-i\omega t}) + it\frac{\lambda^2}{\omega} - \frac{\lambda^2}{\omega^2}(1 - e^{-i\omega t}) \tag{4.357}$$

The phonon correlation function is now

$$U_{ph}(t) = e^{-it(\varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l)} \exp\left[\sum_{\mathbf{q}} \frac{\lambda^2}{\omega_{\mathbf{q}}^2}(1 - e^{-i\omega_{\mathbf{q}} t})\right] \mathrm{Tr}(e^{-\beta H_0} e^{a^\dagger \phi(t)} e^{-a\phi^*(t)})$$

Note that the factor Δ_{kl} has canceled out of this expression. The trace is evaluated using the previous result (4.240). The normalization factor $\exp[\beta\Omega]$ has not been written explicitly in these equations but obviously should be included. The final result is

$$\begin{aligned} U_{ph}(t) &= \exp[-it(\varepsilon_k - \varepsilon_l - \Delta_k + \Delta_l) - \Phi_{kl}(t)] \\ \Phi_{kl}(t) &= \sum_{\mathbf{q}} \frac{(M_{\mathbf{q}k} - M_{\mathbf{q}l})^2}{\omega_{\mathbf{q}}^2}[(N_{\mathbf{q}} + 1)(1 - e^{-i\omega_{\mathbf{q}}t}) + N_{\mathbf{q}}(1 - e^{i\omega_{\mathbf{q}}t})] \end{aligned} \tag{4.358}$$

which is precisely the earlier result in (4.308).

The following physical picture emerges from this derivation. The coupled system of particle and phonon has been sitting in a state, say l, where the phonons have been fluctuating around the equilibrium position. Suddenly the particle is moved to state k by the optical absorption process. From the point of view of the phonon system, it appears that the potential

$$\delta V = \sum_{\mathbf{q}} (M_{\mathbf{q}k} - M_{\mathbf{q}l})(a_{\mathbf{q}} + a_{\mathbf{q}}^\dagger) \tag{4.359}$$

has been suddenly switched on at that time which is taken to be $t = 0$. The transient response in the phonon system is measured by switching off the potential at a later time and measuring the correlations which result. The resulting function of time describes the temporal evolution of the phonon system. Its Fourier transform gives the phonon sideband structure which is observed in the optical spectra. This picture explains the mirror relation between emission and absorption. In absorption the potential which is switched on is (4.359), while in emission, if the particle transition is in the opposite direction, the potential which is suddenly switched on is just the negative of (4.359).

4.3.6. Linked Cluster Expansion

Another way to evaluate the time-ordered exponential operator

$$U(t) = T\left\langle \exp\left[-i\int_0^t dt_1 V(t_1)\right]\right\rangle \tag{4.360}$$

is by the linked cluster expansion, which was discussed in Sec. 3.6. It is sufficient to derive the result for a single-phonon state of wave vector $\mathbf{q}$. The summation over the $\mathbf{q}$ vector in the exponent is easy to do at the end. Consider the case where $V(t)$ is written as

$$V(t) = \lambda A(t) \tag{4.361}$$

$$A(t) = ae^{-i\omega t} + a^\dagger e^{i\omega t} \tag{4.362}$$

The S matrix is expanded directly, and each term is evaluated. The series is presumed to be an exponential series. In the present example, there is only one term which remains after the resummation. The nth-order term is

$$U(t) = \sum_{n=0}^{\infty} (-i)^n U_n(t) \tag{4.363}$$

$$U_n(t) = \frac{\lambda^n}{n!}\int_0^t dt_1 \int_0^t dt_2 \cdots \int_0^t dt_n \langle TA(t_1)\cdots A(t_n)\rangle \tag{4.364}$$

Since $A(t)$ describes the creation or destruction of a phonon, the A operators always exist in pairs. Only terms with n even are nonzero; the others are zero. Since n is even, define $n = 2m$:

$$U(t) = \sum_{m=0}^{\infty} (-i)^{2m} U_{2m}(t) \tag{4.365}$$

$$U_{2m}(t) = \frac{\lambda^{2m}}{(2m)!} \int_0^t dt_1 \int_0^t dt_2 \cdots \int_0^t dt_{2m} \langle TA(t_1) \cdots A(t_{2m}) \rangle \tag{4.366}$$

According to Wick's theorem, the A operators pair up to form phonon Green's function

$$\begin{aligned} \langle TA(t_1)A(t_2) \rangle &= iD(t_1 - t_2) = (N+1)e^{-i\omega|t_1 - t_2|} + Ne^{i\omega|t_1 - t_2|} \\ N &= \frac{1}{e^{\beta\omega} - 1} \end{aligned} \tag{4.367}$$

Each phonon Green's function depends on the difference of two time variables $D(t_i - t_j)$. The time integrals over those two variables define a function

$$\phi(t) = i \int_0^t dt_1 \int_0^t dt_2 D(t_1 - t_2) \tag{4.368}$$

$$= \frac{2}{\omega^2} [(N+1)(1 - e^{-i\omega t}) + N(1 - e^{i\omega t})] \tag{4.369}$$

The mth term has m phonons:

$$\langle TA(t_1) \cdots A(t_{2m}) \rangle = \sum_{\text{all combinations}} D(t_1 - t_j) D(t_j - t_i) \cdots D(t_u - t_{2m})$$

The quantity $U_{2m}(t)$ is just proportional to $\phi(t)^m$, since each combination of time-ordered products gives this same result. The only remaining question is the combinatorial determination of the number of different arrangements. The number of such combinations is

$$\frac{(2m)!}{2^m m!} \tag{4.370}$$

which is obtained in the following way. The first variable, t_1, may be paired with any of the $2m - 1$ other variables. The next time variable, which is t_2 if it was not paired with t_1, is paired with any of the other $2m - 3$ variables. The number of combinations is

$$(2m-1)(2m-3) \cdots 3 \cdot 1 = \frac{(2m)!}{2^m m!} \tag{4.371}$$

The result is

$$U_{2m}(t) = \frac{\lambda^{2m}}{m!} \left[\frac{\phi(t)}{2} \right]^m \tag{4.372}$$

which is summed to obtain the correlation function:

$$U(t) = \exp\left[-\frac{\lambda^2}{2} \phi(t) \right] \tag{4.373}$$

The function $\lambda^2 \phi/2$ is now recognized as the exponential function of time which has already been derived several different ways. It is just the double integral over the phonon Green's

function. The linked cluster expansion is obviously simple, since there is only one distinct linked cluster, which is $\phi(t)$.

Another way to evaluate the Green's function is to leave the particle operators in the S-matrix expansion. There are occasions when it is not immediately apparent that they can be factored out, as was done in (4.328), to give the starting point for (4.346). There are several exactly solvable models which can be evaluated this way. It is useful to learn the technique. Consider the Green's function for a single particle in a band, which is coupled to phonons. The subscript $\mathbf{q}$ is suppressed again in the notation:

$$H = H_0 + V, \quad H_0 = \varepsilon_0 C^\dagger C + \omega a^\dagger a \tag{4.374}$$

$$V = \lambda C^\dagger C(a + a^\dagger) \equiv \lambda C^\dagger C A, \quad A = a + a^\dagger \tag{4.375}$$

$$G(t) = -i\langle TC(t)C^\dagger(0)\rangle = -i\theta(t)\langle e^{iH_0 t}C^\dagger e^{-iHt}C\rangle \tag{4.376}$$

$$G^{(0)}(t) = -i\Theta(t)e^{-i\varepsilon_0 t} \tag{4.377}$$

Since there is only one particle in the band, the creation operator must be to the right, so $t > 0$. The unperturbed Green's function $G^{(0)}(t)$ has a simple form. The interacting Green's function is written, following Chapter 2, as

$$G(t) = -i\langle T\hat{C}(t)U(t)\hat{C}^\dagger(0)\rangle \tag{4.378}$$

The $U(t)$ matrix, or S matrix, is now expanded in an infinite series of terms. This infinite series will generate the linked cluster expansion. However, now the particle operators $C^\dagger$ and C are included, along with the phonon operators a and $a^\dagger$. Again only terms with n even are nonzero, and the $2m$th term is

$$\begin{aligned} G_{2m}(t) = -i\Theta(t)\frac{(-i\lambda)^{2m}}{(2m)!}\int_0^t dt_t \cdots \int_0^1 dt_{2m}\langle T\hat{A}(t_1)\hat{A}(t_2)\cdots\hat{A}(t_{2m})\rangle \\ \times \langle T\hat{C}(t)\hat{C}^\dagger(t_1)\hat{C}(t_1)\cdots\hat{C}^\dagger(t_{2m})\hat{C}(t_{2m})\hat{C}^\dagger(0)\rangle \end{aligned} \tag{4.379}$$

Now there is a time-ordered product of particle operators. At first the time-ordering appears to complicate the evaluation procedure. After all, this time-ordered product of operators,

$$W(t, t_1, \ldots, t_{2m}) = \langle T\hat{C}(t)\hat{C}^\dagger(t_1)\hat{C}(t_1)\cdots\hat{C}^\dagger(t_{2m})\hat{C}(t_{2m})\hat{C}^\dagger(0)\rangle \tag{4.380}$$

must be expanded, according to Wick's theorem, into all possible pairings. However, a careful inspection shows that the time-ordered product has a trivial evaluation. The number operator $C^\dagger C$ is time independent in the interaction representation since H_0 commutes with the number operator,

$$C^\dagger(t_j)C(t_j) = e^{iH_0 t_j}C^\dagger C e^{-iH_0 t_j} = C^\dagger C = n \tag{4.381}$$

so that the time-ordered correlation function just contains a product of number operators:

$$W(t, t_1, \ldots, t_{2m}) = \langle T\hat{C}(t)n^{2m}\hat{C}^\dagger(0)\rangle = \langle T\hat{C}(t)\hat{C}^\dagger(0)\rangle \tag{4.382}$$

The number operator gives unity when operating upon the state with one particle. The W function does not actually depend on any time variable except t. The term remaining, on the

right-hand side, is just proportional to $G^{(0)}(t)$. Since this same factor occurs in each term in the S-matrix expansion it may be factored out of the series:

$$\begin{aligned} G(t) &= \sum_{m=0}^{\infty} G_{2m}(t) \\ G_{2m}(t) &= -iG^{(0)}(t)\frac{(-i\lambda)^{2m}}{(2m)!}\int_0^t dt_1 \cdots \int_0^t dt_{2m}\langle T\hat{A}(t_1)\hat{A}(t_2)\cdots\hat{A}(t_{2m})\rangle \end{aligned} \tag{4.383}$$

What remains in the series is just the same linked cluster expansion which was evaluated above. The exact Green's function is, again, our result which is now very familiar:

$$G(t) = G^{(0)}(t)\ \exp\left[-\frac{\lambda^2}{2}\phi(t)\right] \tag{4.384}$$

The presence of the C operators in the linked cluster expansion did not, in this case, change the result. Usually that does not happen. For most Hamiltonians, the presence of the C operators changes the evaluation of each term in the S-matrix expansion. These cases are often not exactly solvable and certainly are not by the method under discussion.

The usual polaron problem, for one free particle in a band, is described by the Hamiltonian derived in Sec. 1.3:

$$H = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} C_{\mathbf{k}}^{\dagger} C_{\mathbf{k}} + \sum_{\mathbf{q}} [\omega_{\mathbf{q}} a_{\mathbf{q}}^{\dagger} a_{\mathbf{q}} + M_{\mathbf{q}} A_{\mathbf{q}} \rho(\mathbf{q})] \tag{4.385}$$

$$A_{\mathbf{q}} = a_{\mathbf{q}} + a_{-\mathbf{q}}^{\dagger} \tag{4.386}$$

$$\rho(\mathbf{q}) = \sum_{\mathbf{k}} C_{\mathbf{k}+\mathbf{q}}^{\dagger} C_{\mathbf{k}} \tag{4.387}$$

The coupling between the particle and phonons depends on the particle density operator,

$$\rho(\mathbf{q}, t) = \sum_{\mathbf{k}} C_{\mathbf{k}+\mathbf{q}}^{\dagger} C_{\mathbf{k}} e^{it(\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}})} \tag{4.388}$$

which is time dependent. If one were to attempt to solve for the particle Green's function by a linked cluster expansion, the particle part of the correlation function becomes far more complicated. This Green's function is

$$\begin{aligned} G(\mathbf{k}, t) &= -i\Theta(t)\langle e^{iHt} C_{\mathbf{k}} e^{-iHt} C_{\mathbf{k}}^{\dagger}\rangle = -i\Theta(t)\langle e^{iH_0 t} C_{\mathbf{k}} e^{-i(H_0+V)t} C_{\mathbf{k}}^{\dagger}\rangle \\ V &= \sum_{\mathbf{q}} M_{\mathbf{q}} A_{\mathbf{q}} \rho(\mathbf{q}). \end{aligned} \tag{4.389}$$

When evaluated by expanding the S matrix, it produces the $2m$th term of the same general form as (4.379):

$$\begin{aligned} G_{2m}(\mathbf{k}, t) = -i\Theta(t)\frac{(-i)^{2m}}{(2m)!}\int_0^t dt_1 \cdots \int_0^t dt_{2m} \sum_{\mathbf{q}_1} \cdots \sum_{\mathbf{q}_{2m}} M_{\mathbf{q}_1} M_{\mathbf{q}_2} \cdots M_{\mathbf{q}_{2m}} \\ \times \langle T\hat{A}_{\mathbf{q}_1}(t_1) \cdots \hat{A}_{\mathbf{q}_{2m}}(t_{2m})\rangle \langle T\hat{C}(t)\rho(\mathbf{q}_1, t_1) \cdots \rho(\mathbf{q}_{2m}, t_{2m})\hat{C}^{\dagger}(0)\rangle \end{aligned} \tag{4.390}$$

Now the particle part of the correlation function depends on products of the density operators at different times. Usually this function does not have a simple time dependence, and the polaron problem is not exactly solvable. One may still evaluate it by a linked cluster expansion. But this process now becomes an approximate procedure, whereby one evaluates

only a few terms in an infinite series, and the remaining terms are omitted. This approximation will be discussed in Chapter 7.

Van Haeringen (1965) proved the theorem which specifies the most general possible conditions for which (4.389) may be solved exactly by the linked cluster method. An exact solution is obtained whenever $\rho(\mathbf{q}, t)$ has the form

$$\rho(\mathbf{q}, t) = e^{itf(\mathbf{q})}\rho(\mathbf{q}, 0) \tag{4.391}$$

$$\rho(\mathbf{q}, 0) \equiv \rho(\mathbf{q}) = \sum_{\mathbf{k}} C^{\dagger}_{\mathbf{k}+\mathbf{q}} C_{\mathbf{k}} \tag{4.392}$$

Since the general time dependence is given by the kinetic energy difference $f(\mathbf{k}, \mathbf{q}) = \varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}$, his condition is that this difference is independent of $\mathbf{k}$:

$$f(\mathbf{q}) = \varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}} \tag{4.393}$$

If these conditions are met, the exact solution to the Green's function may still be obtained from the linked cluster method. Since the factor $\exp[itf(\mathbf{q})]$ is a c number, and not an operator, it may be removed from the time-ordered product of particle operators:

$$\langle T\hat{C}(t)\rho(\mathbf{q}_1, t_1)\cdots\rho(\mathbf{q}_{2m}, t_{2m})\hat{C}^{\dagger}(0)\rangle = \exp\left[i\sum_{j=1}^{2m} t_j f(\mathbf{q}_j)\right]\langle T\hat{C}(t)\rho(\mathbf{q}_1)\cdots\rho(\mathbf{q}_{2m})\hat{C}^{\dagger}(0)\rangle \tag{4.394}$$

This step leaves, in the time-ordered product, just the time-independent density operators $\rho(\mathbf{q})$. These operators have the same effect that the number operators did previously. They do not change the value of the correlation function. It is exactly equal to

$$\exp\left[i\sum_{j=1}^{2m} t_j f(\mathbf{q}_j)\right]\langle T\hat{C}(t)\hat{C}^{\dagger}(0)\rangle = iG^{(0)}(\mathbf{k}, t)\exp\left[i\sum_{j=1}^{2m} t_j f(\mathbf{q}_j)\right] \tag{4.395}$$

The validity of (4.395) may be shown in the following way. The effect of $\rho(\mathbf{q})$ on $C^{\dagger}_{\mathbf{k}}|0\rangle$ is just to change the particle from the state $\mathbf{k}$ to $\mathbf{k}+\mathbf{q}$:

$$\rho(\mathbf{q})C^{\dagger}_{\mathbf{k}}|0\rangle = \sum_{\mathbf{k}'} C^{\dagger}_{\mathbf{k}'+\mathbf{q}} C_{\mathbf{k}'} C^{\dagger}_{\mathbf{k}}|0\rangle = C^{\dagger}_{\mathbf{k}+\mathbf{q}}|0\rangle \tag{4.396}$$

The effect of a product of such density operators is to change $C^{\dagger}_{\mathbf{k}}|0\rangle$ by a summation of all their $\mathbf{q}$ vectors:

$$\rho(\mathbf{q}_1)\rho(\mathbf{q}_2)\cdots\rho(\mathbf{q}_{2m})C^{\dagger}_{\mathbf{k}}|0\rangle = C^{\dagger}_{\mathbf{Q}}|0\rangle \tag{4.397}$$

$$\mathbf{Q} = \mathbf{k} + \sum_j \mathbf{q}_j = \mathbf{k} \tag{4.398}$$

$$\langle T\hat{C}(t)\rho(\mathbf{q}_1, t_1)\cdots\rho(\mathbf{q}_{2m}, t_{2m})\hat{C}^{\dagger}(0)\rangle = \langle T\hat{C}(t)\hat{C}^{\dagger}(0)\rangle = iG^{(0)}(\mathbf{k}, t)$$

The result requires that the summation over all of the phonon wave vectors is zero. This cancellation is required by the phonon part of the correlation function. When this correlation function is evaluated, by pairing according to Wick's theorem, the pairing forces the $\mathbf{q}$ values to be equal and opposite in sign:

$$\langle TA_{\mathbf{q}_i}(t_i)A_{\mathbf{q}_j}(t_j)\rangle = \delta_{\mathbf{q}_i=-\mathbf{q}_j}\langle TA_{\mathbf{q}_i}(t_i)A_{-\mathbf{q}_i}(t_j)\rangle = i\delta_{\mathbf{q}_i=-\mathbf{q}_j}D(\mathbf{q}_i, t_i - t_j)$$

The summation of all the $\mathbf{q}$ values is certainly zero, since they are paired in sets of $\mathbf{q}_i$ and $-\mathbf{q}_i$. The assertion (4.395) is proven.

The extra time factors $\exp(itf)$ go into the evaluation of the linked cluster expansion. Define a new function:

$$\phi_f(t) = \frac{1}{2}\sum_{\mathbf{q}} M_{\mathbf{q}}^2 \int_0^t dt_1 \int_0^t dt_2 iD(\mathbf{q}, t_1 - t_2) e^{i[t_1 f(\mathbf{q}) + t_2 f(-\mathbf{q})]} \tag{4.399}$$

It is evaluated for the case $f(-\mathbf{q}) = -f(\mathbf{q})$, which is the interesting physical case discussed below:

$$\phi_f(t) = \sum_{\mathbf{q}} M_{\mathbf{q}}^2 \left\{ \frac{-it\omega_{\mathbf{q}}}{\omega_{\mathbf{q}}^2 - f^2} + (N_{\mathbf{q}} + 1)\left[\frac{1 - e^{-it(\omega_{\mathbf{q}} - f)}}{(\omega_{\mathbf{q}} - f)^2}\right] + N_{\mathbf{q}}\left[\frac{1 - e^{-it(\omega_{\mathbf{q}} + f)}}{(\omega_{\mathbf{q}} + f)^2}\right]\right\} \tag{4.400}$$

Terms with $\omega_{\mathbf{q}} - f(\mathbf{q})$ are identical to those with $\omega_{\mathbf{q}} + f(\mathbf{q})$ by the variable change: $\mathbf{q} \to -\mathbf{q}$. The exact result for the Green's function (4.389) is

$$G(\mathbf{k}, t) = -i\Theta(t) e^{-i\varepsilon_{\mathbf{k}} t} e^{-\phi_f(t)} \tag{4.401}$$

Equation (4.401) is the most general form for $G(\mathbf{k}, t)$, which may be exactly solvable, according to van Haeringen.

What kind of systems will obey his condition that $\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}} = f(\mathbf{q})$? One possibility is that the energy is a constant, $\varepsilon_{\mathbf{k}} = \varepsilon_0$ so that $f = 0$. It is this example which has been repeatedly worked throughout this section. However, there is another case which is now recognized as an exactly solvable model. It has the particle moving with a constant velocity $\mathbf{v}$, so that its energy is

$$\varepsilon_{\mathbf{k}} = \mathbf{v} \cdot \mathbf{k} \tag{4.402}$$

$$f(\mathbf{q}) = \mathbf{v} \cdot \mathbf{q} = -f(-\mathbf{q}) \tag{4.403}$$

It should be remarked that the system of equations is not Galilean invariant. Even though the particle is moving with constant velocity, one cannot transform the equations into a system where it is standing still, in a stationary system of phonons. The phonons, particularly the acoustical phonons, have their own velocity and are not invariant under a Galilean transformation. Instead, the phonon frequencies appear Doppler shifted to the particle. The particle Green's function could be evaluated for the case that it was going faster or slower than the acoustical phonons system. Presumably the phonon relaxation around the particle will vary considerably between these two cases. This model is illustrated in the assigned problems. The one-dimensional constant velocity model for polarons was solved by Engelsberg and Varga (1964).

In the polaron problem the particle–phonon coupling produces fluctuations in the number of phonons surrounding the particle. This number fluctuates from time to time and from particle to particle. In the independent boson model, each phonon, in this fluctuating cloud around the particle, is assumed to exist with a probability which is independent of whether other phonons are also simultaneously present. Each phonon is fluctuating independently of the others. If a particle is fixed, after emitting a phonon it is still fixed. The probability of emitting the second phonon is the same as the probability of emitting the first. When the probability of each emission or absorption is the same, then one has the independent boson model. Of course, a particle moving with a constant velocity $\mathbf{v}$ also has its motion unaffected by how many phonons have been emitted. Again the probability of

emitting each phonon is independent of others which may have been emitted, so that all phonon emissions are independent.

A free particle with kinetic energy $\varepsilon_k = k^2/2m$ is not described by the independent boson model. After emission of one phonon with wave vector $\mathbf{q}$, the particle goes to an intermediate state with wave vector $\mathbf{k} + \mathbf{q}$ and energy $(\mathbf{k} + \mathbf{q})^2/2m$. The intermediate state is different from the starting one, so that the emission of the second phonon has a probability that is different from the emission of the first. Each emission has a different probability, so that the probability of n phonon emissions is not just a Poisson distribution,

$$P_n = e^{-\alpha}\frac{\alpha^n}{n!} \tag{4.404}$$

where α is the probability of a single emission. The Poisson distribution is the characteristic zero-temperature distribution of phonon sidebands only when each emission has the same probability, which is independent of the number of other phonons emitted.

The independent boson model has been used widely, with a number of variations. Almbladh and Minnhagen (1978) solved the Fano–Anderson model with phonon coupling to the localized level. A number of related models have been reviewed by Cini and D'Andrea (1988).

4.4. BETHE LATTICE

The terms Bethe lattice and Cayley tree both refer to a type of lattice in which there are no paths which are loops. Each site has z nearest neighbors, where z is an integer greater than one. Since $z = 2$ is a one-dimensional chain, one usually takes $z > 2$. Figure 4.12(a) shows the Bethe lattice for $z = 3$, where a site is at each vertex. Starting from the central point, there are three neighbors. Each of those neighbors also has three neighbors. Of course, each point has the same symmetry.

4.4.1. Electron Green's Function

The Green's function for the Bethe lattice will be calculated for the Hamiltonian with the nearest neighbor hopping term

$$H = -t\sum_{j,\delta} C_j^\dagger C_{j+\delta} - \mu\sum_j C_j^\dagger C_j \tag{4.405}$$

The summation over δ runs over the z neighbors. Spin plays no role in this calculation and will be omitted. The last term contains the chemical potential μ. So how does one calculate the band energy of a lattice without a regular crystal structure? The solution was given by Brinkman and Rice (1970).

As usual, the Hamiltonian is written as $H = H_0 + V$. In this case set $H_0 = -\mu N$. The hopping term is put into V. Define the Green's function $\mathcal{G}(l, ip_n)$ as that needed to advance l

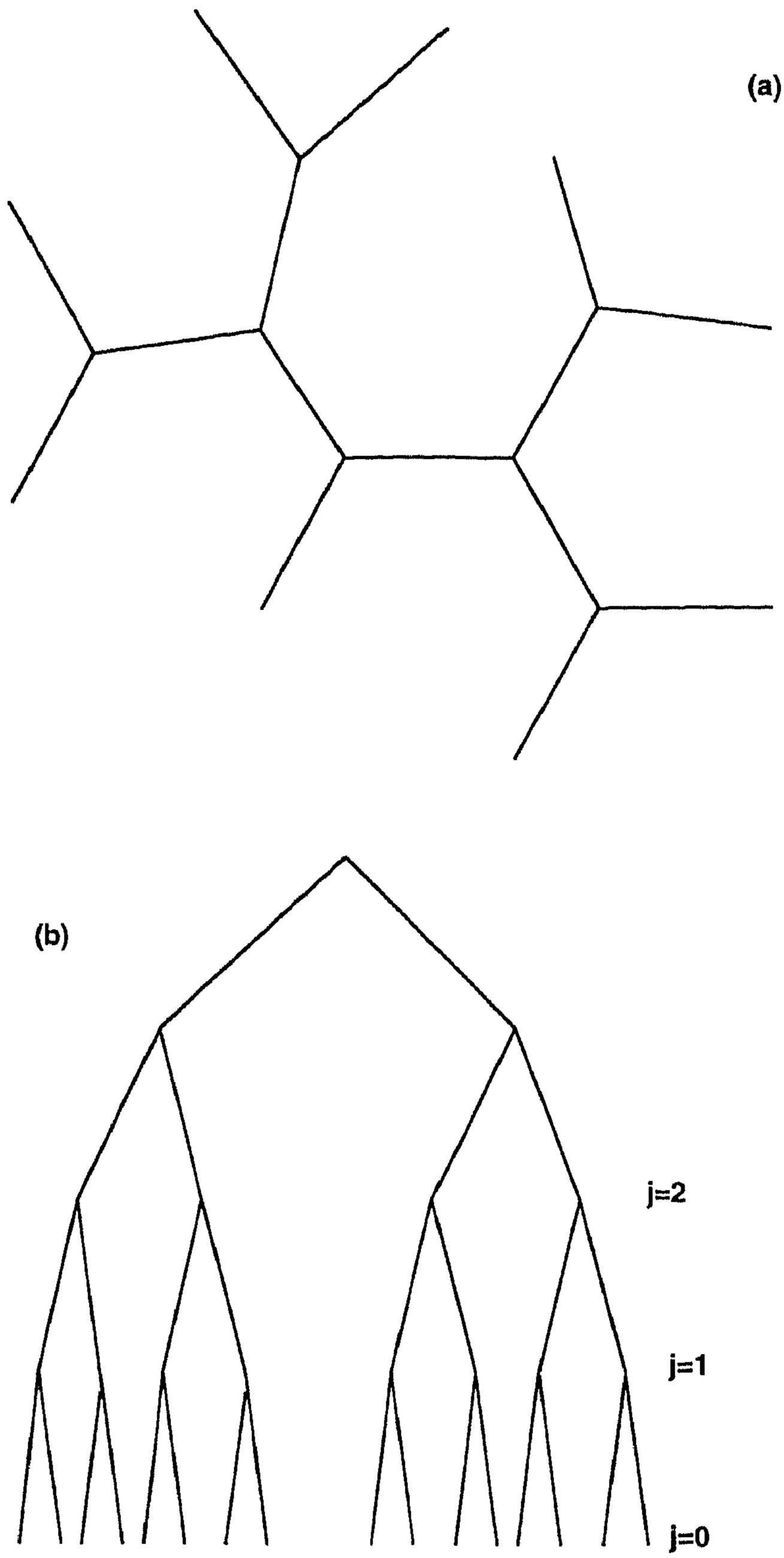


FIGURE 4.12 (a) Bethe lattice for $z = 3$ as usually drawn. (b) The same lattice with a boundary.

steps in the lattice. Since $H_0 = -\mu N$ the noninteracting Green's function for $l = 0$ is given below:

$$\mathcal{G}(l, ip_n) = -\int_0^\beta d\tau e^{ip_n\tau} \langle T_\tau C_{j+l}(\tau) C_j^\dagger(0)\rangle \tag{4.406}$$

$$\mathcal{G}^{(0)}(0, ip_n) = -\int_0^\beta d\tau e^{ip_n\tau} \langle T_\tau C_j(\tau) C_j^\dagger(0)\rangle = \frac{1}{ip_n + \mu} \tag{4.407}$$

Every site has the same Green's function, so there is no need to put on a label such as "j".

The self-energy $\Sigma(0, ip_n)$ is found using perturbation theory. Since the interactions are simple, it is possible to find the exact expression. The self-energy is found from processes where an electron hops to its neighbors and then hops back. Each hop has a matrix element t. The total number of hops in each self-energy is an even number such as $2n$. For $2n$ hops, which return the particle to the original site, the self-energy is denoted as $\Sigma^{(n)}$. The lowest-order self-energy ($n = 1$) is from the process where a particle hops to the neighbors and then hops back to the original site. There are z neighbors, and the intermediate Green's function at the neighboring site must be $\mathscr{G}^{(0)}(0, ip_n)$. The self-energy from double hops is

$$\Sigma^{(1)}(0, ip_n) = \frac{zt^2}{ip_n + \mu} \tag{4.408}$$

The next term in the self-energy comes from the fourth order of perturbation theory, and includes terms where the particle hops twice to a second-nearest neighbor, and then hops back. The first hop has z choices, but the second only has $z - 1$ since the hop back to the starting point has already been included. This term is

$$\Sigma^{(2)}(0, ip_n) = \frac{z(z-1)t^4}{(ip_n + \mu)^3} \tag{4.409}$$

A better way to write the self-energy, including both terms, is

$$\Sigma^{(1+2)}(0, ip_n) = \frac{zt^2}{ip_n + \mu - \dfrac{(z-1)t^2}{ip_n + \mu}} \tag{4.410}$$

$$= \frac{zt^2}{ip_n + \mu}\left[1 + \frac{(z-1)t^2}{(ip_n + \mu)^2} + \left(\frac{(z-1)t^2}{(ip_n + \mu)^2}\right)^2 + \cdots\right] \tag{4.411}$$

The terms in the series in brackets describe processes where, after hopping to the first neighbor, the particle hops to-and-from its neighbors multiple times before hopping back to the original site. It can only hop back to the original site once. When it hops away from the original site the second time, it counts as another self-energy function. Equation (4.411) includes terms which are higher order in the perturbation expansion.

The factor multiplying zt^2 in Eq. (4.410) has the appearance of a Green's function. It is the Green's function for the particle as it hops around. The same Green's function should be in its own denominator. This process leads to a continued fraction

$$\Sigma(0, ip_n) = \cfrac{zt^2}{ip_n + \mu - \cfrac{(z-1)t^2}{ip_n + \mu - \cfrac{(z-1)t^2}{ip_n + \mu - \ddots}}} \tag{4.412}$$

Every denominator is the same. Further terms in the continued fraction describe processes where the particle hops further away in the lattice. The continued fraction contains all forward hops to all sites. This expression is the exact self-energy.

The self-energy (4.410) can be solved in a simple way by defining the self-energy in the denominator as Σ_D

$$\Sigma_D = \frac{(z-1)t^2}{ip_n + \mu - \Sigma_D} \tag{4.413}$$

The continued fraction yields a quadratic equation which is solved easily

$$0 = \Sigma_D^2 - (ip_n + \mu)\Sigma_D + (z-1)t^2 \tag{4.414}$$

$$\Sigma_D = \tfrac{1}{2}\left[ip_n + \mu - \sqrt{(ip_n + \mu)^2 - 4(z-1)t^2}\right] \tag{4.415}$$

$$\Sigma(0, ip_n) = \frac{zt^2}{ip_n + \mu - \Sigma_D} \tag{4.416}$$

$$= \frac{2zt^2}{ip_n + \mu + \sqrt{(ip_n + \mu)^2 - 4(z-1)t^2}} \tag{4.417}$$

$$\mathscr{G}(0, ip_n) = \frac{1}{ip_n + \mu - \Sigma(0, ip_n)} \tag{4.418}$$

Eq. (4.417) is the exact self-energy, and (4.418) is the exact Green's function, for the Bethe lattice in the hopping model.

The self-energy can be used to calculate the density of states. First, recall how it is evaluated for the tight-binding model of a regular crystal. The density of states in d dimensions for a crystal is

$$N(E) = \int \frac{d^d k}{(2\pi)^d} \delta(E - \varepsilon(\mathbf{k})) \tag{4.419}$$

$$\varepsilon(\mathbf{k}) = -t \sum_{\bar{\delta}} e^{i\mathbf{k}\cdot\delta} \tag{4.420}$$

This expression is also just the imaginary part of the retarded real space Green's function $G_{\text{ret}}(\mathbf{r}, E)$ evaluated at $\mathbf{r} = 0$:

$$G_{\text{ret}}(\mathbf{r}, E) = \int \frac{d^d k}{(2\pi)^d} \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{E - \varepsilon(\mathbf{k}) + i\eta} \tag{4.421}$$

$$-2\,\text{Im}\{G(\mathbf{r} = 0, E)\} = 2\pi \int \frac{d^d k}{(2\pi)^d} \delta(E - \varepsilon(\mathbf{k})) \tag{4.422}$$

$$= 2\pi N(E) \tag{4.423}$$

For the Bethe lattice, the above Green's function is in real space at lattice spacing equal to zero. Obtaining the retarded function ($ip_n + \mu \to E + i\eta$), and taking its imaginary part gives

$$\pi N(E) = -\mathrm{Im}[G_{\mathrm{ret}}(0, E)] = -\mathrm{Im}\left[\frac{1}{E + i\delta - \Sigma_{\mathrm{ret}}(0, E - \mu)}\right] \tag{4.424}$$

$$\pi N(E) = \frac{E_s^2\sqrt{E_t^2 - E^2}}{E_s^4 - 4t^2E^2}\Theta(E_t^2 - E^2) \tag{4.425}$$

$$E_s^2 = 2zt^2 \tag{4.426}$$

$$E_t^2 = 4(z-1)t^2 \tag{4.427}$$

It is useful to renormalize the energy by $\Omega = E/(zt)$ and to write the result in dimensionless form

$$\rho(\Omega) = 2\pi t N(E) \tag{4.428}$$

$$= \frac{\sqrt{\Omega_0^2 - \Omega^2}}{1 - \Omega^2}\Theta(\Omega_0^2 - \Omega^2) \tag{4.429}$$

$$\Omega_0^2 = \frac{4(z-1)}{z^2} \tag{4.430}$$

One can verify that for $z = 2$ then $\Omega_0 = 1$ and the density of states is

$$\rho(\Omega) = \frac{\Theta(1 - \Omega^2)}{\sqrt{1 - \Omega^2}} \tag{4.431}$$

This result is identical to that of a particle on a one-dimensional chain with nearest neighbor hopping. Then its energy is $\varepsilon_k = -2t\cos(k)$ and the density of states for this dispersion is the inverse square root function given above. The theory gives the exact result for a chain of atoms, when $z = 2$. For larger values of $z > 2$ then $\Omega_0 < 1$. In these cases the denominator $1 - \Omega^2$ is never zero. The function $\rho(\Omega)$ is shown in Fig. 4.13 for the cases $z = 4$ and $z = 6$. The density of states has a smooth shape over a continuous band of energies. The offsite Green's function is evaluated in the homework problems.

4.4.2. Ising Model

The Ising model for a lattice of spins is derived in Chapter 1. The usual case is to have spin one-half particles at each lattice site. Their z component of spin can point either up or down, which is called $\sigma = \pm 1$. The spins on neighboring sites $\langle ij \rangle$ interact by $-J\sigma_i\sigma_j$. The Ising model with a magnetic field can be solved exactly, analytically, for a one-dimensional chain. In two dimensions it can be solved exactly without a magnetic field, as first done by Onsager. It is interesting that it can be solved exactly for the Bethe lattice. The Hamiltonian is

$$H = -J\sum_{\langle ij \rangle}\sigma_i\sigma_j - h\sum_j \sigma_j \tag{4.432}$$

where $h \equiv \mu H$ is the Zeeman energy in a magnetic field.

The parameter z is the coordination number. It is the number of first neighbors for each site. A related parameter is $r = z - 1$ is called the "branching ratio". When a particle comes to a site, it has r choices of paths to move forward. Here the particles are not moving, but only their spins interact. Figure 4.12(a) shows a Bethe lattice with $r = 2$. The same lattice will be

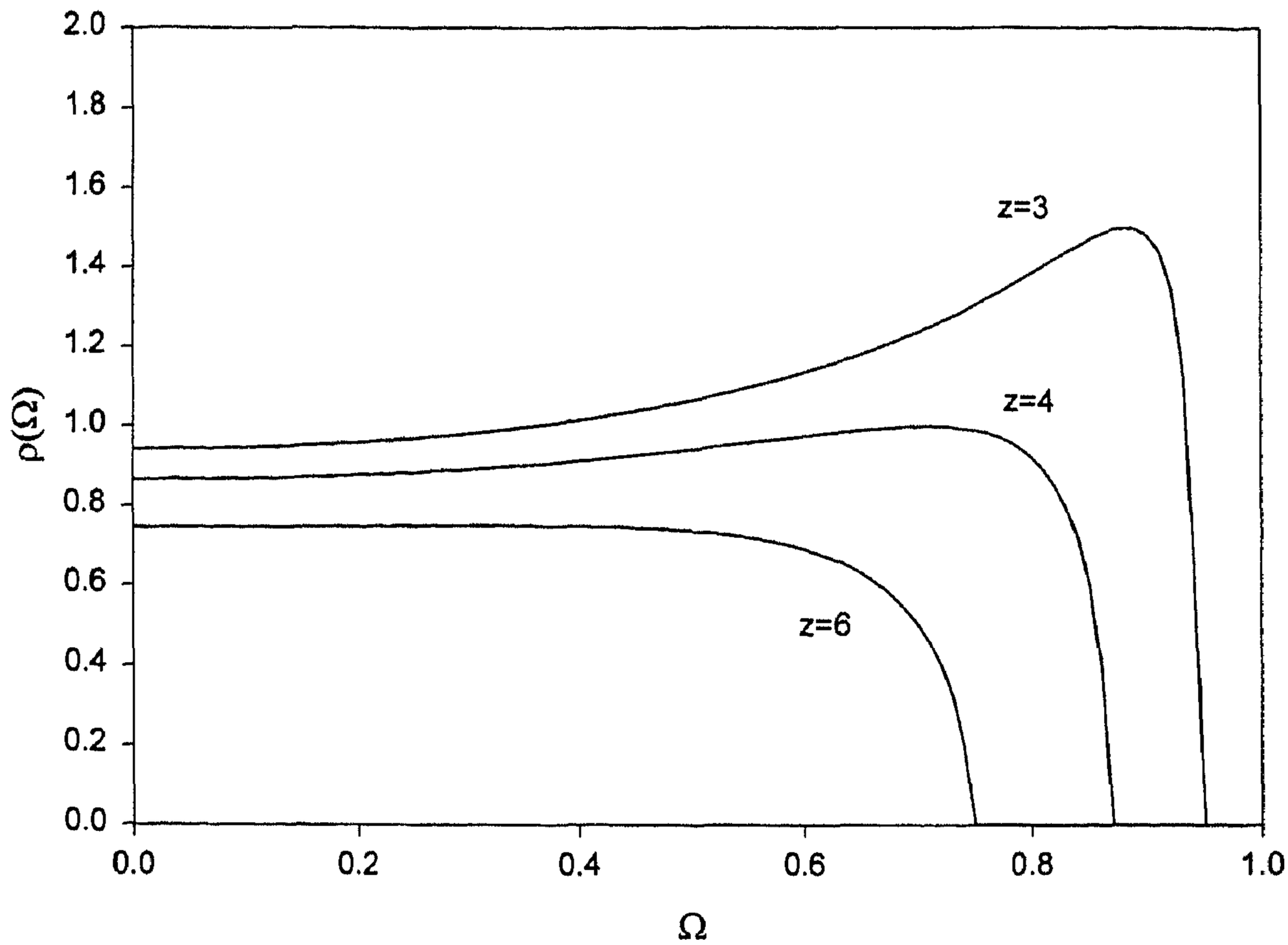


FIGURE 4.13 Normalized density of states for the Bethe lattice for $z = 3$, $z = 4$, and $z = 6$.

drawn in a different way in Fig. 4.12(b). If the lattice is composed of sticks with hinges at the vertices, then if it is picked up at one point, all of the sticks will fall, thereby producing the version of the lattice in Fig. 4.12(b). This version is convenient for solving the Ising model. The bottom of the figure is called the "boundary" of the lattice. The partition function is obtained by averaging the spins at the boundary, and then moving inward row-by-row.

The first step is to average all of the spins in the $j = 0$ row. Each spin in the $j = 1$ row is connected to r spins in the $j = 0$ row. Let σ denote a spin in the $j = 1$ row, and let $(\sigma_1, \sigma_2, \ldots, \sigma_r)$ denote the r spins connected to it in the $j = 0$ row. Averaging just this small complex of spins gives a contribution to the partition function

$$\begin{aligned}\sum_{\sigma_1 \cdots \sigma_r = \pm 1} & \exp\{\beta J \sigma(\sigma_1 + \sigma_2 + \cdots + \sigma_r) + \beta h(\sigma + \sigma_1 + \sigma_2 + \cdots + \sigma_r)\} \\ &= e^{\beta h \sigma}[2\cosh(\beta J \sigma + \beta h)]^r \end{aligned} \tag{4.433}$$

Each spin has an effective interaction with its r neighbors in the lower row of $\sigma[h + J(\sigma_1 + \cdots + \sigma_r)]$. In the nonmagnetic state, the sum over the neighbors will average out to a small number. In the magnetic state, the sum over the neighbors could add up to $\pm r$. It is useful to define the effective magnetic energy h_j which acts upon the spins in row j. A shorthand notation will be that $B_j = \beta h_j$. For the first row there are no spins below, so that $h_0 = h, B_0 = \beta h$. For the next row, Eq. (4.433) is used to define B_1 as

$$A_1 e^{B_1 \sigma} = e^{\beta h \sigma}[2\cosh(B_0 + \beta J \sigma)]^r \tag{4.434}$$

Setting $\sigma = \pm 1$ gives two equations which are solved for the two unknowns (A_1, B_1):

$$A_1 e^{B_1} = e^{\beta h}[2\cosh(B_0 + \beta J)]^r \tag{4.435}$$

$$A_1 e^{-B_1} = e^{-\beta h}[2\cosh(B_0 - \beta J)]^r \tag{4.436}$$

$$B_1 = \beta h + \frac{r}{2}\ln\left[\frac{\cosh(B_0 + \beta J)}{\cosh(B_0 - \beta J)}\right] \tag{4.437}$$

$$A_1 = 2^r[\cosh(B_0 + \beta J)\ \cosh(B_0 - \beta J)]^{r/2} \tag{4.438}$$

The factor of B_0 is inserted instead of βh in the argument of the hyperbolic cosines since the effective field is associated with the spins on the lower row. The factor of βh to the right of the equal sign in (4.437) is from the energy of the spin in row $j = 1$. The formula (4.437) can be manipulated by taking

$$\frac{\cosh(B_0 + \beta J)}{\cosh(B_0 - \beta J)} = \frac{\cosh(B_0)\ \cosh(\beta J) + \sinh(B_0)\ \sin(\beta J)}{\cosh(B_0)\cosh(\beta J) - \sinh(B_0)\ \sin(\beta J)} \tag{4.439}$$

$$= \frac{1 + \tanh(B_0)\ \tanh(\beta J)}{1 - \tanh(B_0)\tanh(\beta J)} \tag{4.440}$$

$$\ln\left[\frac{1+x}{1-x}\right] = 2\ \tanh^{-1} x \tag{4.441}$$

These identities combine to give the result

$$B_1 = \beta h + r\ \tanh^{-1}[\tanh(B_0)\ \tanh(\beta J)] \tag{4.442}$$

The second step is to average the spins in the row $j = 1$ to give the effective field $B_2 = \beta h_2$ in the next row. The procedure is exactly the same. The general recursion relation, as the rows are averaged one-by-one, is

$$B_{j+1} = \beta h + r\ \tanh^{-1}[\tanh(B_j)\ \tanh(\beta J)] \tag{4.443}$$

As the spins are averaged, row-by-row, the effective field $B_j = \beta h_j$ converges to the value in the interior of the Bethe lattice. This bulk value is denoted as B^*. It obeys the self-consistent nonlinear equation

$$B^* = \beta h + r\ \tanh^{-1}[\tanh(B^*)\ \tanh(\beta J)] \tag{4.444}$$

$$\tanh[(B^* - \beta h)/r] = \tanh(B^*)\ \tanh(\beta J) \tag{4.445}$$

The solutions to this equation describe the collective states of the Ising model on the Bethe lattice.

The first example is to assume: (i) zero magnetic field ($h = 0$), and ferromagnetic coupling ($J > 0$). The transition temperature T_c, ($\beta_c = 1/k_B T_c$), is where the ordering begins as one lowers the temperature. At the transition temperature the order parameter B^* is zero,

and it increases in value as the temperature is lowered. It has an infinitesimal value at a temperature which is infinitesimally smaller than T_c. In this case the above equation is

$$B^* = rB^* \tanh(\beta_c J) \tag{4.446}$$

$$\tanh(\beta_c J) = \frac{1}{r} \tag{4.447}$$

$$k_B T_c = \frac{2J}{\ln\left(\frac{r+1}{r-1}\right)} \tag{4.448}$$

$$K_c = \beta_c J = \frac{1}{2}\ln\left(\frac{r+1}{r-1}\right) \tag{4.449}$$

The case for $r = 1$ is a one-dimensional chain. In this case $T_c = 0$ and there is no ordered state at nonzero temperature. For all other branching ratios $r > 1$ there is a well-defined transition temperature.

It is interesting to compare these results of the Bethe lattice with the results for the Ising model on crystalline lattices. What is actually tabulated in Table 4.1 is the quantity K_c for various values of r. Exact results are also shown for crystalline lattices in two and three dimensions (Mahan and Claro, 1977). The lattices are: honeycomb (hc), square (sq), plane-triangular (pt), simple-cubic (sc), body-centered-cubic (bcc) and face-centered-cubic (fcc). The value of K_c for the Bethe lattice is typically below the crystalline result by 10–20%. However, the Bethe lattice has the correct trend that K_c decreases, and T_c increases, as the branching ratio r increases. The crystalline results are analytically known in two dimensions, since the Ising model can be solved exactly. For example, $K_c = \frac{1}{2}\ln(2+\sqrt{3})$ for the hc lattice and $K_c = \frac{1}{2}\ln(1+\sqrt{2})$ for the sq lattice.

Another solution to Eq. (4.444) is to solve for the limit that the temperature goes to zero when the magnetic field is zero. In the limit that $K = \beta J \to \infty$ the equation to solve is

$$\tanh(B^*/r) = \tanh(B^*)\ \tanh(K) \tag{4.450}$$

In the limit that the argument of the hyperbolic tangent function becomes large, its value approaches one. The limit is

$$\lim_{K\to\infty} \tanh(K) = \frac{e^K - e^{-K}}{e^K + e^{-K}} = 1 - 2e^{-2K} + O(e^{-4K}) \tag{4.451}$$

TABLE 4.1 Ising model ferromagnetic transition termperatures ($K_c = J/k_B T_c$). The dimension is d and r is the branching ratio. The column K_c is the exact crystalline result, while $K_c(B)$ is the result from Bethe lattice

d	Crystal	r	K_c	$K_c(B)$
2	hc	2	0.659	0.549
2	sq	3	0.441	0.347
2	pt	5	0.274	0.203
3	sc	5	0.222	0.203
3	bcc	7	0.157	0.144
3	fcc	11	0.102	0.091

so that (4.450) becomes

$$e^{-2B^*/r} = e^{-2B^*} + e^{-2K} \tag{4.452}$$

$$B^* \approx rK \tag{4.453}$$

$$h^* = rJ \tag{4.454}$$

where $B^* = \beta h^*$. The solution to (4.450) is that the effective local field from the r spins in the row below is just $h^* = rJ$. At zero temperature, all of the spins are aligned in the ferromagnetic arrangement.

The derivation at the beginning of the section began by trying to evaluate the partition function. It was used to derive a self-consistent equation for the order parameter B^*. Now return to the task of finding the partition function. The first step is to count the number of spins in the Bethe lattice. The counting is done using Fig. 4.12(b). The top row ($j = L$) has one spin. The next row down has z spins. The next row below that has zr. Successive rows have zr^l. The boundary row has ($j = 0$) has $N_0 = zr^{L-1}$. The number in each row is $N_j = zr^{L-1-j}$ except $N_L = 1$. The total number of spins is $N = \sum N_j = 1 + z(r^L - 1)/(r - 1)$.

The partition function was averaged for a cluster of r spins in the boundary row. The number of such clusters is N_1. So the total partition function after averaging the first row is

$$Z_1 = A_1^{N_1} \exp\left[B_1 \sum_{i=1}^{N_1} \sigma_i\right] \tag{4.455}$$

$$= \mathscr{A}_0^{N_0} \exp\left[B_1 \sum_{i=1}^{N_1} \sigma_i\right] \tag{4.456}$$

$$\mathscr{A}_0 = A_1^{1/r} = 2[\cosh(B_0 + \beta J) \cosh(B_0 - \beta J)]^{1/2} \tag{4.457}$$

This process is repeated for the second row ($j = 1$) which gives the result, including the first two rows, of

$$Z_2 = \mathscr{A}_0^{N_0} \mathscr{A}_1^{N_1} \exp\left[B_2 \sum_{i=1}^{N_2} \sigma_i\right] \tag{4.458}$$

$$\mathscr{A}_1 = 2[\cosh(B_1 + \beta J) \cosh(B_1 - \beta J)]^{1/2} \tag{4.459}$$

This process is repeated for the subsequent rows. The important contribution to the partition function for each row is the factor of $\mathscr{A}_j^{N_j}$. The final partition function is

$$Z = \Pi_j \mathscr{A}_j^{N_j} \approx \mathscr{A}^{*N} \tag{4.460}$$

$$\mathscr{A}^* = 2[\cosh(B^* + \beta J) \cosh(B^* - \beta J)]^{1/2} \tag{4.461}$$

In the last equation, the bulk partition function is assumed to be independent of edge effects. The value of the order parameter for the bulk (B^*) is used in the definition of $\mathscr{A}^*$. That is, assume that B_j rises to its bulk value after a few rows, so that the bulk value is relevent for the

majority of the spin sites. The expression for $\mathscr{A}^*$ can be shortened. The quantity in brackets is (where $K = \beta J$):

$$\begin{aligned}\cosh(B^* + K)\ \cosh(B^* - K) &= \cosh^2(B^*)\ \cosh^2(K) - \sinh^2(B^*)\ \sinh^2(K) \\ &= \cosh^2(B^*)\ \cosh^2(K)[1 - \tanh^2(B^*)\ \tanh^2(K)] \\ &= \cosh^2(B^*)\ \cosh^2(K)[1 - \tanh^2((B^* - \beta h)/r)] \\ &= \frac{\cosh^2(B^*)\ \cosh^2(K)}{\cosh^2[(B^* - \beta h)/r]} \qquad (4.462)\end{aligned}$$

$$\mathscr{A}^* = 2\frac{\cosh(B^*)\cosh(K)}{\cosh[(B^* - \beta h)/r]} \qquad (4.463)$$

where Eq. (4.444) is used to simplify the expression.

The ferromagnetic system has a sharp phase transition only at zero magnetic field ($h = 0$). In this case $B^* = 0$ above the transition temperature, and then $\mathscr{A}^* = 2\cosh(K)$. The expression (4.463) gives the formula in the ordered state by setting $h = 0$. In the case of nonzero magnetic field ($h \neq 0$), the ordering of the spins is gradual as the temperature is lowered. Even at high temperatures, the parameter B^* does not vanish but approaches $B^* \to \beta h$. In this case one should use (4.463) with all of its various factors.

4.5. TOMONAGA MODEL

The Tomonaga model (Tomonaga, 1950) describes a one-dimensional electron gas. The procedure is to examine the Hamiltonian of the one-dimensional electron gas and make some approximations on it. As a consequence of these approximations, the Hamiltonian becomes exactly solvable. The one-dimensional electron gas is not exactly solvable but only an approximate version of it.

The important physics is the recognition that the excitations of the electron gas are approximate bosons, although the elementary particles, electrons, are fermions. The excitations involve two-particle states, for example, moving an electron from one state to another. The wave function of the two fermion states has boson properties. The Tomonaga model assumes that the excitations are exactly bosons, which is the important approximation.

The model has been useful in several kinds of problems. First, there are organic solids such as TTF–TCNQ whose conductivity is thought to be largely one dimensional (see Heeger, 1977). The Tomonaga model has played a role in the interpretation of electrical conductivity in these materials (see Luther and Emery, 1974). Second, in impurity problems, or X-ray absorption problems, the response of the electron gas to the central impulse can be factored into spherical harmonics associated with different angular momentum states l. Each angular momentum channel l then becomes a one-dimensional electron gas to which one may apply the Tomonaga model. Recently, semiconductor nanotechnology permits the construction of semiconductor channels which act as one-dimensional conductors. The Tomonaga model is used in the theory of these systems. Single wall carbon nanotubes are another one-dimensional conductor.

4.5.1. Tomonaga Model

The original model of Tomonaga (1950) discusses the following Hamiltonian for the one-dimensional interacting electron gas:

$$H = v_F \sum_{ks} |k| a_{ks}^\dagger a_{ks} + \frac{1}{2L} \sum_k V_k \rho(k)\rho(-k) \tag{4.464}$$

$$\rho(k) = \sum_{ps} a_{p-k/2,s}^\dagger a_{p+k/2,s} \tag{4.465}$$

The system has length L, and v_F is the Fermi velocity of the particles, which are assumed to have a linear dispersion relation. The label $s = \pm 1$ denotes spin, and $\rho(k)$ is the electron density operator. The electron–electron interaction term V_k will be specified below. It is not $4\pi e^2/k^2$, which is dimensionally incorrect in one dimension, since V_k has units of Joule-meter. Dimensional analysis suggests the form $V_k \propto e^2(k_F/k)^n$, where n is any exponent. The summation over k states may be turned into integrals by the usual transformation as $L \to \infty$:

$$\sum_k f(k) = \frac{L}{2\pi} \int dk f(k) \tag{4.466}$$

The basic step in the Tomonaga model is to divide the density operator into two terms:

$$\rho_1(k) = \sum_{p>0,s} a_{p-k/2,s}^\dagger a_{p+k/2,s} \tag{4.467}$$

$$\rho_2(k) = \sum_{p<0,s} a_{p-k/2,s}^\dagger a_{p+k/2,s} \tag{4.468}$$

$$\rho(k) = \rho_1(k) + \rho_2(k) \tag{4.469}$$

The density operator $\rho(k)$ commutes with any other density operator $\rho(k')$. However, the two parts ρ_1 and ρ_2 do not commute with the same parts for other wave vectors. Examine the commutation relations:

$$\begin{aligned}[\rho_1(k), \rho_1(k')] &= \sum_{s,s'} \sum_{p,p'>0} [a_{p-k/2,s}^\dagger a_{p+k/2,s}, a_{p'-k'/2,s'}^\dagger a_{p'+k'/2,s'}] \\ &= \sum_{s,p>0} [a_{p-k/2,s}^\dagger a_{p+k'+k/2,s} \Theta(p + k/2 + k'/2) \\ &\quad - a_{p-k'-k/2,s}^\dagger a_{p+k/2,s} \Theta(p - k/2 - k'/2)]\end{aligned} \tag{4.470}$$

An important special case is $k' = -k$

$$[\rho_1(k), \rho_1(-k)] = \sum_{s,p>0} [n_{p-k/2,s} - n_{p+k/2,s}] = \sum_s \sum_{-k/2 \le p \le k/2} n_{p,s} \tag{4.471}$$

The right-hand side shows that the commutation relations depend on the operator n_{ps} over a range of p values. The operator n_{ps} is replaced by its average in the ground state of the free-particle system.

$$\sum_s \sum_{-k/2 \le p \le k/2} n_{ps} = 2 \sum_{-k/2 \le p \le k/2} \Theta(k_F - |p|) = \begin{cases} 2(kL/2\pi), & k < 2k_F \\ 2k_F L/\pi, & k > 2k_F \end{cases}$$

and the commutation relations (4.471) can be written for $k < 2k_F$ as

$$[\rho_1(k), \rho_1(-k)] = \left(\frac{kL}{\pi}\right) \tag{4.472}$$

$$[\rho_2(k), \rho_2(-k)] = -\left(\frac{kL}{\pi}\right) \tag{4.473}$$

$$[\rho_1(k), \rho_2(-k)] = 0 \tag{4.474}$$

The analogous results are included for the other commutators, which can be derived in the same fashion. The Tomonaga model assumes that these density operators obey the exact commutation relations of

$$\begin{aligned} [\rho_1(k), \rho_1(-k') &= \delta_{k,k'}\left(\frac{kL}{\pi}\right) \\ [\rho_2(k), \rho_2(-k')] &= -\delta_{k,k'}\left(\frac{kL}{\pi}\right) \\ [\rho_1(k), \rho_2(-k')] &= 0 \end{aligned} \tag{4.475}$$

These relations are the central approximation of the Tomonaga model. The commutation relations are not exact, since the commutators give operators, as in (4.470). However, these results are obtained when taking the expectation value of the exact commutation relations. For example, in (4.470)

$$\begin{aligned} \langle[\rho_1(k), \rho_1(k')]\rangle = \sum_{s,p>0} [&\langle a^\dagger_{p-k/2,s} a_{p+k'+k/2,s}\rangle \Theta(p + k/2 + k'/2) \\ &- \langle a^\dagger_{p-k'-k/2,s} a_{p+k/2,s}\rangle \Theta(p - k/2 - k'/2)] \end{aligned} \tag{4.476}$$

In the right-hand side, the averages are zero unless $k' = -k$, so that

$$\langle[\rho_1(k), \rho_1(-k')]\rangle = \delta_{k,k'} \sum_{s,p>0} [\langle n_{p-k/2}\rangle - \langle n_{p+k/2}\rangle] = 2\delta_{k,k'} \sum_{-k/2\le p\le k/2} \langle n_{ps}\rangle \tag{4.477}$$

Although the commutation relations (4.475) are not exact, the expectation values of these commutators are given exactly. The approximation is not a very bad one.

It is convenient to express the density operators $\rho_j(\pm k)$ in terms of creation and destruction operators. This step is done so that the creation operators are dimensionless and

the commutation relations (4.475) are obeyed. The creation and destruction operators are for bosons. These definitions are given below, where the symbol k is always positive:

$$\begin{aligned}
\rho_1(k) &= b_k\sqrt{\frac{kL}{\pi}} \\
\rho_1(-k) &= b_k^\dagger\sqrt{\frac{kL}{\pi}} \\
\rho_2(k) &= b_{-k}^\dagger\sqrt{\frac{kL}{\pi}} \\
\rho_2(-k) &= b_{-k}\sqrt{\frac{kL}{\pi}}
\end{aligned} \tag{4.478}$$

$$[b_k, b_{k'}^\dagger] = \delta_{k,k'} \tag{4.479}$$

When k is positive $\rho_1(k) \propto b_k$, and when k is negative $\rho_1(k) \propto b_{-k}^\dagger$. The operators ρ_1 always commute with ρ_2. The choice (4.478) does satisfy the approximate commutation relations (4.475).

The second term in the Hamiltonian (4.465) may be written in terms of these boson operators:

$$\frac{1}{2L}\sum_k V_k\rho(k)\rho(-k) = \sum_k \bar{V}_k(b_k + b_{-k}^\dagger)(b_k^\dagger + b_{-k}) \tag{4.480}$$

$$\bar{V}_k = \frac{|k|V_k}{2\pi} \tag{4.481}$$

The electron–electron interaction term has been recast into an interaction between the boson excitations of the electron gas.

The first term in (4.465) is the particle kinetic energy. It requires some additional work in order to express it in terms of boson coordinates. It is not immediately obvious how to express $a_k^\dagger a_k$ in terms of the new boson operators. When faced with this predicament, it is useful to examine the commutation relations of this operator. The objective is to find a boson representation of the kinetic energy operator which reproduces the commutation relations. If this cannot be done exactly, at least try to find a good approximation. The commutator algebra completely specifies the excitation spectrum of the system, so that the excitations are adequately described by operators with accurate commutation relations.

Call the kinetic energy term H_0. Its commutator with $\rho_1(k)$ is

$$\begin{aligned}
[\rho_1(k), H_0] &= v_F \sum_{s,p>0}\sum_{s'k'} |k'|[a_{p-k/2,s}^\dagger a_{p+k/2,s}, a_{k's'}^\dagger a_{k's'}] \\
&= v_F \sum_{s,p>0} a_{p-k/2,s}^\dagger a_{p+k/2,s}(|p+k/2| - |p-k/2|)
\end{aligned} \tag{4.482}$$

$$|p+k/2| - |p-k/2| = \begin{cases} k & \text{if } p > k/2 \\ 2p & \text{if } p < k/2 \end{cases} \tag{4.483}$$

For small values of k, then $p > k/2$ over most of the p summation. In this case the above commutator is approximately given by

$$[\rho_1(k), H_0] = v_F k \sum_{s,p>0} a^\dagger_{p-k/2,s} a_{p+k/2,s} = v_F k \rho_1(k) \tag{4.484}$$

The above is a desirable form for the commutator, since the right-hand side is also proportional to $\rho_1(k)$. With the boson representations (4.475) and (4.478) for $\rho_1(k)$, the approximate commutation relation (4.484) is

$$[b_k, H_0] = k v_F b_k = \omega_k b_k \tag{4.485}$$

Of course, the same result would be given by the choice of $H_0 = \sum_k \omega_k b_k^\dagger b_k$. Next consider the commutator of H_0 with ρ_2. The same approximation in this case leads to

$$[\rho_2(k), H_0] = -\omega_k \rho_2(k) \tag{4.486}$$

Both of these approximate commutators are satisfied with the following choice for H_0:

$$H_0 = \sum_k \omega_k b_k^\dagger b_k \tag{4.487}$$

$$H = \sum_k \{\omega_k b_k^\dagger b_k + \bar{V}_k (b_k + b_{-k}^\dagger)(b_k^\dagger + b_{-k})\} \tag{4.488}$$

The one-dimensional electron gas (4.465) has been recast into the boson Hamiltonian (4.488). The latter is exactly solvable, as will soon be shown. The Tomonaga model (4.488) has been derived from (4.465) with several key approximations on commutation relations. The form (4.488) is a description of the boson excitations of the electron gas.

Equation (4.488) may be solved exactly by a variety of techniques. Probably the easiest method is to change to a coordinate representation for the boson operators:

$$Q_k = \frac{1}{\sqrt{2\omega_k}} (b_k + b_{-k}^\dagger) \tag{4.489}$$

$$P_k = i\sqrt{\frac{\omega_k}{2}} (b_k^\dagger - b_{-k}) \tag{4.490}$$

$$[Q_k, P_{k'}] = i\delta_{k,k'} \tag{4.491}$$

In this representation the Hamiltonian is written as

$$H_0 = \tfrac{1}{2} \sum_k (P_{-k} P_k + \omega_k^2 Q_k Q_{-k}) \tag{4.492}$$

$$H = \tfrac{1}{2} \sum_k (P_{-k} P_k + E_k^2 Q_k Q_{-k}) \tag{4.493}$$

$$E_k^2 = \omega_k^2 + 4\omega_k \bar{V}_k \tag{4.494}$$

The new eigenfrequencies are E_k. Now change back to a new set of boson normal mode operators, which are normalized to the new eigenfrequencies.

$$Q_k = \frac{1}{\sqrt{2E_k}}(\alpha_k + \alpha_{-k}^\dagger) \tag{4.495}$$

$$P_k = i\sqrt{\frac{E_k}{2}}(\alpha_k^\dagger - \alpha_{-k}) \tag{4.496}$$

$$[\alpha_k, \alpha_{k'}^\dagger] = i\delta_{k,k'} \tag{4.497}$$

$$H = \sum_k E_k(\alpha_k^\dagger \alpha_k + \tfrac{1}{2}) \tag{4.498}$$

These series of steps may be summarized by the observation that the boson operators are changed in the following way:

$$b_k + b_{-k}^\dagger = \sqrt{\frac{\omega_k}{E_k}}(\alpha_k + \alpha_{-k}^\dagger) \tag{4.499}$$

$$b_k^\dagger - b_{-k} = \sqrt{\frac{E_k}{\omega_k}}(\alpha_k^\dagger - \alpha_{-k}) \tag{4.500}$$

These transformations are useful for other problems.

The Hamiltonian of the one-dimensional electron gas (4.465) has been solved approximately. Only the excitation spectrum has been obtained. Some of these excitations are fluctuations in the density operator $\rho(k)$. Very similar results to the Tomonaga model are obtained by writing an equation of motion for the density operator and solving it approximately. This approach is used in Chapter 5.

So far the form of the interaction potential V_k has not been specified. In fact, physicists choose a variety of forms for this interaction to suit their problem. The units of V_k are the same as v_F: Joule-meter ($\hbar v_F$ is J-m). One possible choice is to take $V_k \propto e^2 = \text{constant} = V_0$. The energy spectrum is just altered by having the Fermi velocity increased:

$$E_k = \bar{v}_F k \tag{4.501}$$

$$\bar{v}_F = \sqrt{v_F\left(v_F + \frac{2}{\pi}V_0\right)} \tag{4.502}$$

The constant V_0 is assumed to be positive, since it describes interactions between electrons. The interactions increase the velocity of the acoustic plasmon.

Another possible choice is to take $V_k = 2/3(e^2k_F^2/k^2)$. This choice leads to long-wavelength modes with a constant frequency, which is the plasma frequency:

$$E_k = \sqrt{k^2 v_F^2 + \omega_p^2} \tag{4.503}$$

$$\omega_p^2 = 4\omega_k \bar{V}_k = \frac{4\pi e^2 n_0}{m} \tag{4.504}$$

where $n_0 = k_F^3/3\pi^2$ is not the electron density in one dimension, but is a collection of constants. In the electron gas, there are two different types of excitations. One is the plasma modes at long wavelength, and the other is the electron–hole excitations at shorter wavelength. The latter are probably best described by the choice $V_k = V_0$.

4.5.2. Spin Waves

The Hamiltonian (4.464) of the one-dimensional electron gas has other collective excitations besides the density oscillations which were discussed above. These other excitations have the character of spin waves, or magnons. Overhauser (1965) has shown that the excitation spectrum is completely described by the sum of these two types of excitations: density oscillations and spin waves. This feature of one dimension does not apply to three dimensions. The density oscillations are the excitations which occur when there are external perturbations such as electric fields. The spin waves respond to magnetic perturbations and contribute to the spin susceptibility.

The spin waves are described by the operators

$$\sigma(k) = \sigma_1(k) + \sigma_2(k) \tag{4.505}$$

$$\sigma_1(k) = \sum_{p>0,s} s a^\dagger_{p-k/2,s} a_{p+k/2,s} \tag{4.506}$$

$$\sigma_2(k) = \sum_{p<0,s} s a^\dagger_{p-k/2,s} a_{p+k/2,s} \tag{4.507}$$

where the spin index is $s = \pm 1$ for $\uparrow, \downarrow$. The nature of the spin wave excitations is shown in Fig. 4.14. The spin-up and spin-down densities have opposite variations, so there is no net change in the particle density. There is a variation in $\rho_\uparrow - \rho_\downarrow$

$$\rho_s(k) = \sum_p a^\dagger_{p-k/2,s} a_{p+k/2,s} \tag{4.508}$$

$$\rho = \rho_\uparrow + \rho_\downarrow \tag{4.509}$$

$$\sigma = \rho_\uparrow - \rho_\downarrow \tag{4.510}$$

The spin operators are examined in the same fashion used for the density operators. The commutation relations are found among these operators and between them and the density operators. Some typical results are

$$[\sigma_1(k), \sigma_1(-k')] = \delta_{k,k'}\left(\frac{kL}{\pi}\right) \tag{4.511}$$

$$[\sigma_2(k), \sigma_2(-k')] = -\delta_{k,k'}\left(\frac{kL}{\pi}\right) \tag{4.512}$$

$$[\sigma_1(k), \sigma_2(k')] = 0 \tag{4.513}$$

$$[\sigma_i(k), \rho_j(-k')] = 0 \qquad (i, j = 1, 2) \tag{4.514}$$

The commutator $[\sigma_i(k), \rho_j(k)]$ contains one factor of s, and the term $s = 1$ cancels $s = -1$. This cancellation occurs when the two spin states are occupied with equal probability and the system is not magnetic. The spin operators commute with the density operators and so

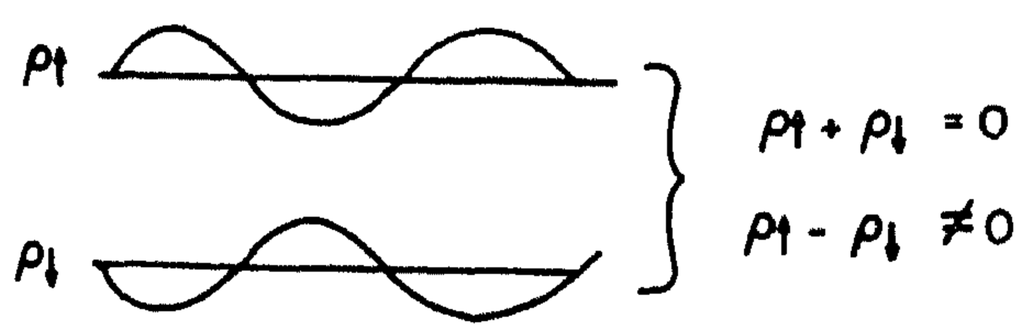


FIGURE 4.14 Spin-up and spin-down charge densities.

describe an independent set of excitations. These excitations can be represented by a new set of creation and destruction operators, which for $k > 0$ are

$$\begin{aligned}\sigma_1(k) &= c_k\sqrt{\frac{kL}{\pi}}\\ \sigma_1(-k) &= c_k^\dagger\sqrt{\frac{kL}{\pi}}\\ \sigma_2(k) &= c_{-k}^\dagger\sqrt{\frac{kL}{\pi}}\\ \sigma_2(-k) &= c_{-k}\sqrt{\frac{kL}{\pi}}\end{aligned} \tag{4.515}$$

$$[c_k, c_{k'}^\dagger] = \delta_{k,k'} \tag{4.516}$$

$$[c_k, b_{k'}^\dagger] = 0 \tag{4.517}$$

The next step is to examine the commutation relation of $\sigma_j(k)$ with the Hamiltonian (4.465), which will establish the energy spectrum of these spin wave operators. They commute with the second term in (4.465), from electron–electron interactions, since they commute with the density operators. The commutator with the kinetic energy term H_0 is:

$$[\sigma_1(k), H_0] = v_F \sum_{s,p>0} s a_{p-k/2,s}^\dagger a_{p+k/2,s}[|p+k/2| - |p-k/2|] \tag{4.518}$$

$$\approx v_F k \sigma_1(k) \tag{4.519}$$

The commutator is evaluated using the same approximation to get (4.484). Exactly the same result is obtained by representing the spin wave part of H_0 by $\sum_k \omega_k c_k^\dagger c_k$. The spin wave part of the Hamiltonian is

$$H_{sw} = \sum_k \omega_k c_k^\dagger c_k \tag{4.520}$$

$$H = \sum_k \{\omega_k b_k^\dagger b_k + \omega_k c_k^\dagger c_k + \bar{V}_k(b_k + b_{-k}^\dagger)(b_k^\dagger + b_{-k})\} \tag{4.521}$$

$$H = \sum_k \{E_k \alpha_k^\dagger \alpha_k + \omega_k c_k^\dagger c_k\} \tag{4.522}$$

The density operator parts in (4.488), (4.522), and (4.498) are combined with the spin wave parts to give the total Hamiltonian H for the excitation spectra of the one-dimensional electron gas. The original model of Tomonaga actually described a spinless electron gas. For spin one-half systems, the two possible spin orientations lead to another type of independent excitation which are called spin waves. The total Hamiltonian (4.522) has the density and spin wave excitations decoupled.

The original Hamiltonian (4.465) did not contain any terms which would cause interactions between spin waves; there were no terms of the type $\sigma(k)\sigma(-k)$. The spin wave excitation spectrum is unchanged by electron–electron interactions, at least in the Tomonaga model.

The spin wave part of the excitation spectrum can be used to derive the Pauli spin susceptibility. The starting point for this calculation is (3.458):

$$\chi(k, i\omega) = -\int_0^\beta d\tau e^{i\omega\tau} \langle T_\tau \sigma(k, \tau)\sigma(-k, 0)\rangle \tag{4.523}$$

In the Tomonaga model, the correlation function may be evaluated exactly by using the operator representation (4.515):

$$\sigma(-k, 0) = \sigma_1(-k) + \sigma_2(-k) = \left(\frac{kL}{\pi}\right)(c_k^\dagger + c_{-k}) \tag{4.524}$$

$$\sigma(k, \tau) = \left(\frac{kL}{\pi}\right)(c_{-k}^\dagger e^{\tau\omega_k} + c_k e^{-\tau\omega_k}) \tag{4.525}$$

The τ dependence is determined by H_{sw}. The further steps in the evaluation of the correlation function are identical to the derivation of the unperturbed phonon Green's function in (3.76):

$$\chi(k, i\omega) = \left(\frac{|k|L}{\pi}\right)\frac{2\omega_k}{(i\omega)^2 - \omega_k^2} \tag{4.526}$$

$$\chi_{\text{ret}}(k, \omega) = \left(\frac{|k|L}{\pi}\right)\frac{2\omega_k}{\omega^2 - \omega_k^2 + i2\omega\delta} \tag{4.527}$$

The retarded correlation function is found from the analytical continuation $i\omega \to \omega + i\delta$.

The susceptibility is found to be proportional to the length L of the electron gas. This dependence on L is correct, since the susceptibility is the total magnetization M divided by the magnetic field, and the total magnetization is indeed proportional to the size of the system. A more meaningful quantity would be the magnetization per unit volume, which is the above result divided by L. The susceptibility demonstrates a resonance phenomenon, so that it is singular whenever the external perturbations (k, ω) exactly match those of the excitation spectrum $\omega = \omega_k = kv_F$.

4.5.3. Luttinger Model

A model proposed by Luttinger (1963) is a slight variation on the Tomonaga model. It has the advantage of being exactly solvable, with fewer approximations, yet is identical to the Tomonaga model in some of its essential properties. The basic feature of the Luttinger model is that the system has two types of fermions. One has an energy spectrum given by $\varepsilon_k = kv_F$, while the other has an energy spectrum given by $\varepsilon_k = -kv_F$. They are shown by the solid and dashed lines in Fig. 4.15(a). There is an infinite number of each kind of particle, since the occupied energy states stretch to negative infinity.

In the Tomonaga model (4.465) it is assumed the energy spectrum is as shown in Fig. 4.15(b). The particles have a linear dispersion relation, but the same kind of particle is represented throughout the band of states.

The two kinds of fermions in the Luttinger model are denoted by the operators $a_{1,k,s}$ and $a_{2,k,s}$, where the subscript 1 or 2 designates the particle. The two bands are quite independent, so the two fermion operators anticommute:

$$\{a_{i,k,s}, a_{j,k',s'}^\dagger\} = \delta_{ij}\delta_{k,k'}\delta_{ss'} \tag{4.528}$$

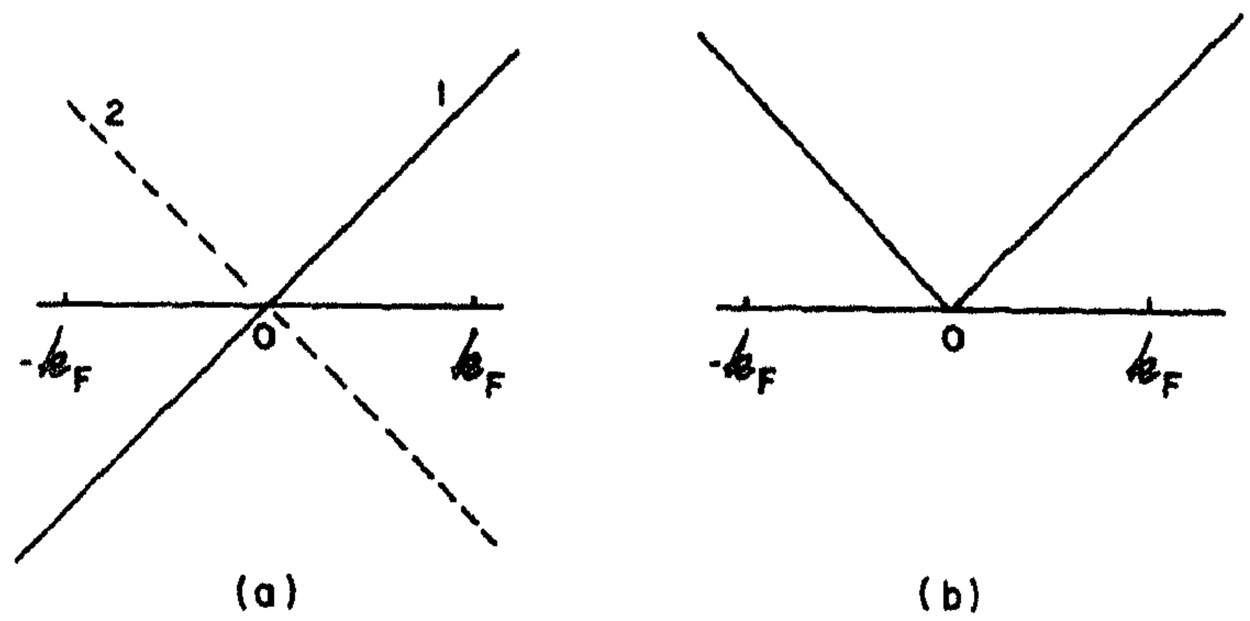


FIGURE 4.15 (a) The Luttinger model has two distinct particles, with separate energy bands. (b) The Tomonaga model has one particle, whose energy band is $v_F|k|$.

The operators $\rho_i(k)$ and $\sigma_j(k)$ are defined as in the Tomonaga model ($p > 0$):

$$\rho_i(p) = \sum_{ks} a^\dagger_{i,k+p,s} a_{i,k,s} \tag{4.529}$$

$$\rho_i(-p) = \sum_{ks} a^\dagger_{i,k,s} a_{i,k+p,s} = \rho_i(p)^\dagger \tag{4.530}$$

$$\sigma_i(p) = \sum_{ks} s a^\dagger_{i,k+p,s} a_{i,k,s} \tag{4.531}$$

$$\sigma_i(-p) = \sum_{ks} a^\dagger_{i,k,s} a_{i,k+p,s} = \sigma_i(p)^\dagger \tag{4.532}$$

The advantage of the Luttinger model is that it has the same kind of commutation relations as found for the Tomonaga model. However, they are valid for all p, whereas they were valid only for $p < 2k_F$ in the Tomonaga model:

$$[\rho_1(-p), \rho_1(p')] = \delta_{p,p'}\left(\frac{pL}{\pi}\right) \tag{4.533}$$

$$[\rho_2(p), \rho_2(-p')] = \delta_{p,p'}\left(\frac{pL}{\pi}\right) \tag{4.534}$$

$$[\rho_1(p), \rho_2(p')] = 0 \tag{4.535}$$

$$[\sigma_1(-p), \sigma_1(p')] = \delta_{p,p'}\left(\frac{pL}{\pi}\right) \tag{4.536}$$

$$[\sigma_2(-p), \sigma_2(p')] = \delta_{p,p'}\left(\frac{pL}{\pi}\right) \tag{4.537}$$

$$[\sigma_1(p), \sigma_2(p')] = 0 \tag{4.538}$$

$$[\sigma_i(p), \rho_j(p')] = 0 \tag{4.539}$$

These commutation relations depend, in an important way, on the assumption that there is an infinite number of negative-energy particles. For example, the first commutator is

$$[\rho_1(-p), \rho_1(p')] = 2\delta_{p,p'} \sum_k (n_{2,k} - n_{1,k+p}) \tag{4.540}$$

The factor of 2 comes from the summation over the two spin configurations $s = \pm 1$. For a finite number of particles, each summation over particle number would just give the number of 1-particles N_1,

$$N_1 = 2\sum_k n_{1,k} = 2\sum_k n_{1,k+p} \tag{4.541}$$

and the commutator would be zero. However, when there is an infinite number of particles in negative-energy states, a nonzero result is obtained. For a finite band, the difference

$$\sum_k (n_{1,k} - n_{1,k+p}) \tag{4.542}$$

equals $pL/2\pi$ at the top end of the band, but it equals the negative of this at the bottom end of the band, so that there is no net difference. For a semi-infinite band, there is no bottom contribution, so only the top difference is counted.

The kinetic energy term in the Luttinger model is

$$H_0 = v_F \sum_{ks} k(a^\dagger_{1,k,s} a_{1,k,s} - a^\dagger_{2,k,s} a_{2,k,s}) \tag{4.543}$$

H_0 has the exact commutation relations with the operators ($p > 0$)

$$[H_0, \rho_1(p)] = v_F p \rho_1(p) \qquad [H_0, \rho_2(p)] = -v_F p \rho_2(p) \tag{4.544}$$

$$[H_0, \sigma_1(p)] = v_F p \sigma_1(p) \qquad [H_0, \sigma_2(p)] = -v_F p \sigma_2(p) \tag{4.545}$$

The kinetic energy term is exactly represented by the operator

$$H_0 = \frac{\pi v_F}{L} \sum_{p>0} [\rho_1(p)\rho_1(-p) + \rho_2(-p)\rho_2(p) + \sigma_1(p)\sigma_1(-p) + \sigma_2(-p)\sigma_2(p)] \tag{4.546}$$

In the Tomonaga model, the boson approximation applies only for excitation with small k. This restriction is removed in the Luttinger model. The transformation to boson operators is

$$\begin{aligned}
\rho_1(-p) &= b_{1p}\sqrt{\frac{pL}{\pi}}, & \rho_1(p) &= b^\dagger_{1p}\sqrt{\frac{pL}{\pi}} \\
\rho_2(-p) &= b^\dagger_{2,-p}\sqrt{\frac{pL}{\pi}}, & \rho_2(p) &= b_{2,-p}\sqrt{\frac{pL}{\pi}} \\
\sigma_1(-p) &= c_{1p}\sqrt{\frac{pL}{\pi}}, & \sigma_1(p) &= c^\dagger_{1p}\sqrt{\frac{pL}{\pi}} \\
\sigma_2(-p) &= c^\dagger_{2,-p}\sqrt{\frac{pL}{\pi}}, & \sigma_2(p) &= c_{2,-p}\sqrt{\frac{pL}{\pi}}
\end{aligned} \tag{4.547}$$

The Hamiltonian is now

$$H_0 = \sum_{p>0} p v_F [b^\dagger_{1p} b_{1p} + b^\dagger_{2,-p} b_{2,-p} + c^\dagger_{1p} c_{1p} + c^\dagger_{2,-p} c_{2,-p}] \tag{4.548}$$

The operator $\rho_1(p)$ for $p > 0$ takes a particle from state k and puts it into $p + k$. This operation will make an electron–hole pair when $k < k_F$ and $p + k > k_F$. The summation over all such electron–hole pairs is represented by the boson creation operator $b^\dagger_{1p}$. For particle 2, the Fermi "surface" is at the negative wave vector $-k_F$. Electron–hole pairs are made mostly at negative wave vectors. The operator $\rho_2(-p) = \sum_k a^\dagger_{2k} a_{2,k+p}$ for $p > 0$ creates these

bosons, since it takes an electron from the occupied state $k_F < k + p$ to the unoccupied state $k < -k_F$, where k is negative.

Various kinds of interaction terms may be added to the Luttinger model. Those which arise from electron–electron interactions are expressed as the product of four fermion operators, or two density operators. These Hamiltonians are exactly solvable, since they describes linear coupling between two harmonic oscillator systems.

The Luttinger model has the advantage of being exactly solvable. Of course, one could add other terms which might render it no longer exactly solvable. The disadvantage of the model is that it is unphysical, since it contains the infinite reservoir of negative-energy particles.

4.5.4. Single-Particle Properties

Some of the most interesting applications of the Tomonaga–Luttinger models are concerned with single-particle properties of the electron gas. An important quantity is the occupation number $n_{i,k,s} = \langle a^\dagger_{i,k,s} a_{i,k,s} \rangle$, in the interacting system. A more ambitious calculation would be the one-particle Green's function

$$G_{i,s}(k, t) = -i\langle T a_{i,k,s}(t) a^\dagger_{i,k,s}(0) \rangle \tag{4.549}$$

To obtain these quantities requires a representation of the single fermion operator $a_{i,k,s}$ in terms of boson operators. The discussion follows Mattis and Lieb (1965) and Luther and Peschel (1974).

The representation of the single-fermion operator in terms of bosons is found, as always, by examining the commutation relations. A representation of $a_{i,k,s}$ is satisfactory if it obeys all the proper commutation relations with the other operators. The first step is to Fourier-transform into a real-space representation:

$$\Psi_{is}(x) = \frac{1}{\sqrt{L}} \sum_k e^{ikx} a_{i,k,s} \tag{4.550}$$

$$\Psi^\dagger_{is}(x) = \frac{1}{\sqrt{L}} \sum_k e^{-ikx} a^\dagger_{i,k,s} \tag{4.551}$$

The advantage of this representation becomes clear when considering the commutator of $\Psi_{is}(x)$ with the density operators. This discussion uses the Luttinger form of the Tomonaga model. Typical commutators are

$$[\Psi_{is}(x), \rho_j(p)] = \delta_{ij} e^{ipx} \Psi_{is}(x) \tag{4.552}$$

$$[\Psi_{is}(x), \sigma_j(p)] = \delta_{ij} s e^{ipx} \Psi_{is}(x) \tag{4.553}$$

which is derived in the following way:

$$\begin{aligned}
[\Psi_{is}(x), \rho_j(p)] &= \frac{1}{\sqrt{L}} \sum_{kk's'} e^{ikx} [a_{i,k,s}, a^\dagger_{j,k'+p,s'} a_{j,k',s'}] \\
&= \frac{\delta_{i,j}}{\sqrt{L}} \sum_{kk's'} e^{ikx} a_{j,k',s'} \delta_{ss'} \delta_{k=k'+p} \\
&= \frac{\delta_{i,j}}{\sqrt{L}} e^{ipx} \sum_{k'} e^{ik'x} a_{i,k',s} = \delta_{i,j} e^{ipx} \Psi_{is}(x)
\end{aligned} \tag{4.554}$$

The commutator $[\Psi_{is}(x), \rho_j(p)]$ has a simple form, since it is just proportional to $\Psi_{is}(x)$. The solution would be simpler if the commutator were a constant or even proportional to a density operator. It is not, so the solution of (4.552) is more complicated. One possible solution has the form

$$\Psi_{1s}(x) = F_1(x)\ \exp[J_1(x)] \tag{4.555}$$

$$J_1(x) = -\frac{\pi}{L}\sum_{p>0}\frac{1}{p}\{e^{-ipx}[\rho_1(p) + s\sigma_1(p)] - e^{ipx}[\rho_1(-p) + s\sigma(-p)]\} \tag{4.556}$$

The prefactor $F_1(x)$ can be a function of x but is a c number in the sense that it must commute with both $\rho_1(p)$ and $\sigma_1(p)$. Next show that this choice does satisfy Eq. (4.552)

$$\begin{aligned}[\Psi_{1s}, \rho_1] &= F_1(e^{J_1}\rho_1 - \rho_1 e^{J_1}) = F_1(e^{J_1}\rho_1 e^{-J_1} - 1)e^{J_1} \\ &= F_1[J_1, \rho_1]e^{J_1} = [J_1, \rho_1]\Psi_{1s}\end{aligned} \tag{4.557}$$

The last line is valid only when the commutator $[J_1, \rho_1]$ is a c number which commutes with the operator J_1. It does for the $J_1(x)$ in (4.556):

$$[J_1, \rho_1(p)] = \frac{\pi}{L}\sum_k \frac{1}{k}e^{ikx}[\rho_1(-k), \rho_1(p)] = e^{ipx} \tag{4.558}$$

The next observation is that the factor

$$\rho_1(p) + s\sigma_1(p) = \sum_{k,s'}(1 + ss')a^\dagger_{1,k+p,s'}a_{1,k,s'} = 2\sum_k a^\dagger_{1,k+p,s}a_{1,k} \tag{4.559}$$

since the factor $(1 + ss') = 0$ unless $s = s'$, and then it is 2. The equations can be condensed by introducing the notation of a spin-dependent density operator ($p > 0$):

$$\rho_{is}(p) = \sum_{\text{all } k} a^\dagger_{i,k+p,s}a_{i,k,s} \tag{4.560}$$

$$\rho_{is}(-p) = \sum_{\text{all } k} a^\dagger_{i,k,s}a_{i,k+p,s} \tag{4.561}$$

$$\rho_i(p) = \sum_s \rho_{is}(p) \tag{4.562}$$

$$\sigma_i(p) = \sum_s s\rho_{is}(p) \tag{4.563}$$

$$J_{1s}(x) = -\frac{2\pi}{L}\sum_{\text{all } p}\frac{e^{-ipx}}{p}\rho_{1s}(p) \tag{4.564}$$

$$\Psi_{1s}(x) = F_1(x)\ \exp[J_{1s}(x)] \tag{4.565}$$

The spin-dependent density operators can be represented by boson operators similar to (4.547) with an additional spin subscript.

The form of $\Psi_{1s}(x)$ in (4.556) is a solution to the commutator equation (4.552). Unfortunately this solution has some undesirable properties which will force a modification. The need for changes in $\Psi_{is}(x)$ may be understood by examining the form of the operator $\langle\Psi^\dagger_{1s}(x)\Psi_{1s}(x')\rangle$ for the noninteracting electron system, which is the Luttinger model with just

the Hamiltonian H_0 in (4.543). At zero temperature, the noninteracting system has the feature that the momentum distributions for particles 1 and 2 have the form

$$n_{1,k,s} = \theta(k_F - k) \tag{4.566}$$

$$n_{2,k,s} = \theta(k_F + k) \tag{4.567}$$

This fact can be used to evaluate the correlation function $\langle \Psi^\dagger_{1s}(x)\Psi_{1s}(x')\rangle$ by using the inverse of the transformation (4.550):

$$\langle \Psi^\dagger_{1s}(x)\Psi_{1s}(x')\rangle = \frac{1}{L}\sum_{kk'} e^{-i(kx-k'x')}\langle a^\dagger_{1ks}a_{1k's}\rangle \tag{4.568}$$

$$= \frac{1}{L}\sum_{kk'} e^{-i(kx-k'x')}\delta_{kk'}n_{1,k,s} \tag{4.569}$$

$$= \int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-ik(x-x')}\theta(k_F - k) \tag{4.570}$$

$$\langle \Psi^\dagger_{1s}(x)\Psi_{1s}(x')\rangle = -\frac{e^{-ik_F(x-x')}}{2\pi i(x - x' + i\eta)} \tag{4.571}$$

$$\langle \Psi^\dagger_{2s}(x)\Psi_{2s}(x')\rangle = \frac{e^{ik_F(x-x')}}{2\pi i(x - x' - i\eta)} \tag{4.572}$$

where the factors of $\pm i\eta$ are added to aid convergence at infinity.

The objective in choosing the representation (4.565) is to make the result for $\langle \Psi^\dagger_{1s}(x)\Psi_{1s}(x')\rangle$ be like (4.571) for the noninteracting Luttinger model. A method of doing this was suggested by Luther and Peschel (1974). It uses a limiting process, where the wave function $\Psi_{1s}(x)$ contains a parameter α, and the limit $\alpha \to 0$ is taken at the end of the calculation. Including the parameter α, the position space operators are represented as

$$\Psi_{1s}(x) = \frac{1}{\sqrt{2\pi\alpha}}\exp[ik_F x + J_{1s}(\alpha, x)] \tag{4.573}$$

$$\Psi_{2s}(x) = \frac{1}{\sqrt{2\pi\alpha}}\exp[-ik_F x - J_{2s}(\alpha, x)] \tag{4.574}$$

$$J_{is}(\alpha, x) = -\frac{2\pi}{L}\sum_{k>0}\frac{e^{-\alpha k/2}}{k}[e^{-ikx}\rho_{is}(k) - e^{ikx}\rho_{is}(-k)] \tag{4.575}$$

$$= -J_{is}(\alpha, x)^\dagger \tag{4.576}$$

This new form for $J_{is}(\alpha, x)$ may be expressed in terms of the boson operators:

$$J_{1s}(\alpha, x) = \sum_{p>0} e^{-\alpha p/2}\sqrt{\frac{2\pi}{pL}}(b_{1s,p}e^{ipx} - b^\dagger_{1s,p}e^{-ipx}) \tag{4.577}$$

$$J_{2s}(\alpha, x) = \sum_{p>0} e^{-\alpha p/2}\sqrt{\frac{2\pi}{pL}}(b^\dagger_{2s,-p}e^{ipx} - b_{2s,-p}e^{-ipx}) \tag{4.578}$$

The exponential factor $J_{1s}(\alpha, x)$ has the same form as (4.556) in the limit where $\alpha \to 0$, so the commutator (4.552) is obeyed in this limit. The prefactor $1/\sqrt{2\pi\alpha}$ is explained below. Now consider the evaluation of the quantity

$$\langle \Psi_{1s}^\dagger(x)\Psi_{1s}(x')\rangle = \frac{1}{2\pi\alpha} e^{-ik_F(x-x')} \langle e^{-J_{1s}(\alpha,x)} e^{J_{1s}(\alpha,x')}\rangle \tag{4.579}$$

$$= \frac{1}{2\pi\alpha} e^{-ik_F(x-x')} \Pi_{k>0} \left\langle \exp\left[e^{-\alpha k/2}\sqrt{\frac{2\pi}{kL}}(e^{-ikx} b_{1s,k}^\dagger - e^{ikx} b_{1s,k})\right]\right.$$

$$\left. \times \exp\left[e^{-\alpha k/2}\sqrt{\frac{2\pi}{kL}}(e^{ikx'} b_{1s,k} - e^{-ikx'} b_{1s,k}^\dagger)\right]\right\rangle \tag{4.580}$$

The right-hand side of this expression is an average of exponential functions of boson operators. These expressions are evaluated in Sec. 4.3.2. Each exponent is separated by using the Feynman theorem $\exp(A+B) = \exp(A)\exp(B)\exp(-[A,B]/2)$. Then the factors are commuted until all the destruction operators are on the right:

$$\langle \Psi_{1s}^\dagger(x)\Psi_{1s}(x')\rangle = \frac{1}{2\pi\alpha} e^{-ik_F(x-x')-\phi_0(x-x')} \Pi_{k>0} \left\langle \exp\left[e^{-\alpha k/2}\sqrt{\frac{2\pi}{kL}} b_{1s,k}^\dagger (e^{-ikx} - e^{-ikx'})\right]\right.$$

$$\left. \times \exp\left[e^{-\alpha k/2}\sqrt{\frac{2\pi}{kL}} b_{1s,k} (e^{ikx'} - e^{ikx})\right]\right\rangle$$

$$\phi_0(x) = \frac{2\pi}{L}\sum_{k>0}\frac{e^{-\alpha k}}{k}(1-e^{ikx}) = \int_0^\infty \frac{dk}{k} e^{-\alpha k}(1-e^{ikx}) \tag{4.582}$$

At zero temperature, the quantity in the final brackets gives unity, which gives the following prediction for the noninteracting electron gas:

$$\langle \Psi_{1s}^\dagger(x)\Psi_{1s}(x')\rangle = \frac{1}{2\pi\alpha} e^{-ik_F(x-x')-\phi_0(x-x')} \tag{4.583}$$

The expression for $\phi_0(x)$ has the form of an infrared divergence as discussed in Sec. 9.3. Expand the exponential $\exp(ikx)$ and integrate term by term. The factor $\exp(-\alpha k)$ ensures the convergence of these integrals, which is the primary role played by α:

$$\phi_0(x) = -\sum_{l=0}^\infty \frac{(ix)^l}{l!}\int_0^\infty dk k^{l-1} e^{-\alpha k} = -\sum_{l=0}^\infty \frac{(ix)^l}{l\alpha^l}$$

$$= \ln\left(1 - \frac{ix}{\alpha}\right) \tag{4.584}$$

$$e^{-\phi_0(x)} = \frac{1}{1 - ix/\alpha} \tag{4.585}$$

The series for $\phi_0(x)$ is recognized as a logarithm, which gives the final result

$$\langle \Psi_{1s}^\dagger(x)\Psi_{1s}(x')\rangle = \frac{1}{2\pi\alpha} e^{-ik_F(x-x')} \frac{1}{1 - i(x-x')/\alpha} \tag{4.586}$$

$$= \frac{1}{2\pi} e^{-ik_F(x-x')} \frac{1}{\alpha - i(x-x')} \tag{4.587}$$

The limit $\alpha \to 0$ does indeed recover the noninteracting value (4.571). The parameter α becomes the convergence factor η.

It is easy to check that the factor $\langle \Psi_{2s}^\dagger(x)\Psi_{2s}(x')\rangle$ is also given correctly. The representation (4.575) for $\Psi_{1s}(x)$ reproduces the commutation relation with the density operators and also gives the correct ground state momentum distribution for the noninteracting system. All these results, of course, apply in the limit where $\alpha \to 0$. The commutator of $\Psi_{1s}(x)$ with H_0 is also given correctly, since the latter is expressed in terms of the density operators, which have the correct commutators. The representation (4.575) is suitable for the single-fermion operators.

This representation can be used to calculate many interesting properties of the Luttinger model. For example, the electron Green's function is

$$\begin{aligned} G_{1s}(x-x', t) &= -i\langle T\Psi_{1s}(x, t)\Psi_{1s}^\dagger(x', 0)\rangle \\ &= -i\theta(t)\langle e^{iH_0t}\Psi_{1s}(x)e^{-iH_0t}\Psi_{1s}^\dagger(x')\rangle \\ &\quad + i\theta(t)\langle \Psi_{1s}^\dagger(x')e^{iH_0t}\Psi_{1s}(x)e^{-iH_0t}\rangle \end{aligned} \tag{4.588}$$

The correlation functions can be evaluated at zero temperature, using the same steps which led to (4.581). The time dependence of $\Psi_{1s}(x, t)$ for the noninteracting Hamiltonian is

$$\Psi_{1s}(x, t) = \frac{1}{\sqrt{2\pi\alpha}} \exp[ik_Fx + J_{1s}(\alpha, x, t)] \tag{4.589}$$

$$\begin{aligned} J_{1s}(\alpha, x, t) &= \sum_{k>0} e^{-\alpha k/2}\sqrt{\frac{2\pi}{kL}}(e^{i(kx-\omega_k t)}b_{1s,k} - e^{-i(kx-\omega_k t)}b_{1s,k}^\dagger) \\ J_{1s}(\alpha, x, t) &= J_{1s}(\alpha, x - v_Ft) \end{aligned} \tag{4.590}$$

The time dependence of $\Psi_{1s}(x, t) = \Psi_{1s}(x - v_Ft)$ merely changes the factor x in $J_{1s}(\alpha, x)$ to $x - v_Ft$. This rather trivial change makes it possible to use the previous result for $\langle \Psi_{1s}^\dagger(x)\Psi_{1s}(x')\rangle$ to evaluate $\langle \Psi_{1s}^\dagger(x')\Psi_{1s}(x, t)\rangle$:

$$\begin{aligned} \langle \Psi_{1s}^\dagger(x')\Psi_{1s}(x, t)\rangle &= \frac{1}{2\pi\alpha} \exp[ik_F(x-x') - \phi_0(x' - x + v_Ft)] \\ &= \frac{e^{ik_F(x-x')}}{2\pi i(x - x' - v_Ft - i\alpha)} \end{aligned} \tag{4.591}$$

$$\begin{aligned} \langle \Psi_{1s}(x, t)\Psi_{1s}^\dagger(x')\rangle &= \frac{1}{2\pi\alpha} \exp[ik_F(x-x') - \phi_0^*(x' - x + v_Ft)] \\ &= -\frac{e^{ik_F(x-x')}}{2\pi i(x - x' - v_Ft + i\alpha)} \end{aligned} \tag{4.592}$$

The factor $\langle \Psi_{1s}(x, t)\Psi_{1s}^\dagger(x')\rangle$ has just the Hermitian conjugate of ϕ_0. The Green's function for the noninteracting system is easily obtained ($\alpha \to 0$):

$$G_{1s}(x, t) = \frac{e^{ik_Fx}}{2\pi}\left[\frac{\theta(t)}{x - v_Ft + i\alpha} + \frac{\theta(-t)}{x - v_Ft - i\alpha}\right] \tag{4.593}$$

This equation can be Fourier-transformed to obtain the Green's function in the wave vector representation:

$$G_{1s}(k,t) = \int_{-\infty}^{\infty} dx e^{-ikx} G_{1s}(x,t) \tag{4.594}$$

$$= -ie^{-iv_F t(k-k_F)}[\Theta(t)\Theta(k-k_F) - \Theta(-t)\Theta(k_F-k) \tag{4.595}$$

This result is the same $G(k, t)$ which is obtain in the fermion representation:

$$G_{1s}(k,t) = -i\langle T a_{i,k,s}(t) a^{\dagger}_{1,k,s}(0)\rangle \tag{4.596}$$

where the energy has been normalized to the Fermi energy: $\varepsilon_k = v_F(k - k_F)$. The correct result for $G_{1s}(k, t)$ again illustrates that the boson representation (4.575) for the single-particle operators will faithfully reproduce the results obtained directly from the fermion representation. The virtue of the Boson representation is that more difficult problems can be solved. In particular, interaction terms can be added to the Hamiltonian. Exact expressions can be found for Green's functions, or other correlation functions, although they are usually difficult to evaluate analytically.

4.5.5. Interacting System of Spinless Fermions

An exact solution can be obtained for various correlation functions, even for the interacting electron gas in one dimension. First solve for the occupation number. This solution relies upon the representation of the single-particle operators which was developed in the prior subsection. The Hamiltonian in this part is taken to be the Luttinger model for spinless fermions (Mattis and Lieb, 1965):

$$H = \sum_{p>0}[\omega_p(b^{\dagger}_{1,p}b_{1,p} + b^{\dagger}_{2,-p}b_{2,-p}) + V_p(b^{\dagger}_{1,p}b^{\dagger}_{2,-p} + b_{2,-p}b_{1,p})] \tag{4.597}$$

The interaction term comes from particle–particle interactions between the two types of fermions. Other interaction terms could be considered.

The first step in the solution is to learn the method of diagonalizing this Hamiltonian. There are several ways to do this, and all give the same result. A canonical transformation is used to obtain a new set of boson operators α_p, β_p, which are defined as

$$b_{1,p} = \beta_p \cosh(\lambda_p) - \alpha^{\dagger}_p \sinh(\lambda_p) \tag{4.598}$$

$$b^{\dagger}_{1,p} = \beta^{\dagger}_p \cosh(\lambda_p) - \alpha_p \sinh(\lambda_p) \tag{4.599}$$

$$b_{2,-p} = \alpha_p \cosh(\lambda_p) - \beta^{\dagger}_p \sinh(\lambda_p) \tag{4.600}$$

$$b^{\dagger}_{2,-p} = \alpha^{\dagger}_p \cosh(\lambda_p) - \beta_p \sinh(\lambda_p) \tag{4.601}$$

$$[b_{1,p}, b^{\dagger}_{1,p}] = [\beta_p, \beta^{\dagger}_p] \cosh^2(\lambda_p) + [\alpha^{\dagger}_p, \alpha_p] \sinh^2(\lambda_p) \tag{4.602}$$

$$= \cosh^2(\lambda_p) - \sinh^2(\lambda_p) = 1 \tag{4.603}$$

$$[\alpha_p, \beta^{\dagger}_p] = 0, \qquad [\alpha_p, \alpha^{\dagger}_p] = 1 \tag{4.604}$$

The various commutation relations are still obeyed in this new representation. The parameter λ_p is chosen so that the Hamiltonian (4.597) is diagonalized. It is first written out in terms of the transformed operators:

$$H = \sum_{p>0}[(\beta_p^\dagger\beta_p + \alpha_p^\dagger\alpha_p)\{[\cosh^2(\lambda_p) + \sinh^2(\lambda_p)]\omega_p - 2V_p \sinh(\lambda_p) \cosh(\lambda_p)\}$$
$$+ (\beta_p^\dagger\alpha_p^\dagger + \alpha_p\beta_p)\{[\cosh^2(\lambda_p) + \sinh^2(\lambda_p)]V_p - 2\omega_p \sinh(\lambda_p) \cosh(\lambda_p)\}]$$

Since these are boson operators, the ordering of terms such as $\alpha\beta = \beta\alpha$ does not matter. The zero-point motion terms are ignored. Two combinations of hyperbolic functions seem to occur:

$$\cosh^2(\lambda_p) + \sinh^2(\lambda_p) = \cosh(2\lambda_p) \tag{4.605}$$

$$2\sinh(\lambda_p) \cosh(\lambda_p) = \sinh(2\lambda_p) \tag{4.606}$$

The Hamiltonian is diagonalized by setting to zero the coefficient of the term $(\beta_p^\dagger\alpha_p^\dagger + \alpha_p\beta_p)$. This step gives $\tanh(2\lambda_p) = V_p/\omega_p$, so that the diagonalized Hamiltonian is

$$H = \sum_{P>0} E_p(\beta_p^\dagger\beta_p + \alpha_p^\dagger\alpha_p) \tag{4.607}$$

$$E_p = \sqrt{\omega_p^2 - V_p^2} \tag{4.608}$$

$$\cosh(2\lambda_p) = \frac{\omega_p}{E_p} \tag{4.609}$$

The transformation to the new operators is used to evaluate the properties of the interacting system. The α_p and β_p operators refer to the actual boson normal modes in the interacting system. The ground state of the system is the vacuum of α_p and β_p particles; i.e., $\alpha_p|0\rangle = 0$, $\beta_p|0\rangle = 0$. These are the same set of normal modes in the Tomonaga model (4.498).

Consider the evaluation of the fermion occupation number, which is given in (4.581) as the ground state expectation value of the operator combination:

$$\langle\Psi_1^\dagger(x)\Psi_1(x')\rangle = \frac{1}{2\pi\alpha} \exp[-ik_F(x-x') - \phi_0(x-x')] \tag{4.610}$$

$$e^{-\phi_0(x)} = \langle e^{-J_1(\alpha,x)}e^{J_1(\alpha,0)}\rangle \tag{4.611}$$

where $J_1(\alpha, x)$ is given in (4.577). The ground state of the system must be the particle vacuum of the bosons with excitation energy E_p in (4.608), since these are the normal modes. The $J_1(\alpha, x)$ operator must be expressed in the α_p and β_p representation. The transformation (4.600) produces a redefined operator form

$$J_1(\alpha, x) = \sum_{p>0} e^{-\alpha p/2}\sqrt{\frac{2\pi}{pL}}\{e^{ipx}[\beta_p \cosh(\lambda_p) - \alpha_p^\dagger \sinh(\lambda_p)]$$
$$- e^{-ipx}[\beta_p^\dagger \cosh(\lambda_p) - \alpha_p \sinh(\lambda_p)]\} \tag{4.612}$$

It contains operators of both types α_p and β_p. These operators are independent, since they each describe an independent Boson system. Each of these boson systems are averaged independently. The ground state average gives

$$\phi_0(x) = \phi_a(x) + \phi_b(x) \tag{4.613}$$

$$e^{-\phi_a(x)} = \langle e^{-J_a(x)} e^{J_a(x)} \rangle \tag{4.614}$$

$$e^{-\phi_b(x)} = \langle e^{-J_b(x)} e^{J_b(x)} \rangle \tag{4.615}$$

$$J_a(x) = \sum_{p>0} e^{-\alpha p/2} \sqrt{\frac{2\pi}{pL}} \sinh(\lambda_p)(e^{-ipx}\alpha_p - e^{ipx}\alpha_p^\dagger) \tag{4.616}$$

$$J_b(x) = \sum_{p>0} e^{-\alpha p/2} \sqrt{\frac{2\pi}{pL}} \cosh(\lambda_p)(e^{ipx}\beta_p - e^{-ipx}\beta_p^\dagger) \tag{4.617}$$

The separate averages for $\phi_a(x)$ and $\phi_b(x)$ are similar to those found earlier in (4.582). The average for $\phi_b(x)$ is identical to the earlier average for $\phi_0(x)$, except for the extra kernel $\cosh(\lambda_p)$. The average for $\phi_a(x)$ also contains a unique kernel $\sinh(\lambda_p)$ and has $x \to -x$. By analogy with (4.582), at zero temperature

$$\phi_b(x) = \frac{2\pi}{L} \sum_{p>0} \frac{e^{-\alpha p}}{p} \cosh^2(\lambda_p)(1 - e^{ipx}) \tag{4.618}$$

$$\phi_a(x) = \frac{2\pi}{L} \sum_{p>0} \frac{e^{-\alpha p}}{p} \sinh^2(\lambda_p)(1 - e^{-ipx}) \tag{4.619}$$

The result for $\phi_b(x)$ is manipulated by replacing $\cosh^2(\lambda_p)$ by its equivalent $1 + \sinh^2(\lambda_p)$. The term with "1" is identical to $\phi_0(x)$

$$\phi_b(x) = \phi_0(x) + \frac{2\pi}{L} \sum_{p>0} \frac{e^{-\alpha p}}{p} \sinh^2(\lambda_p)(1 - e^{ipx}) \tag{4.620}$$

so that

$$\begin{aligned} \phi_a + \phi_b &= \phi_0 + \phi_s \\ \phi_s(x) &= \frac{2\pi}{L} \sum_{p>0} \frac{e^{-\alpha p}}{p} \sinh^2(\lambda_p)[(1 - e^{-ipx}) + (1 - e^{ipx})] \end{aligned} \tag{4.621}$$

$$\langle \Psi_1^\dagger(x)\Psi_1(x') \rangle = -\frac{e^{-ik_F(x-x')}}{2\pi i(x - x' + i\alpha)} e^{-\phi_s(x-x')} \tag{4.622}$$

The effect of the interactions on the electron gas is contained in the exponential factor $\exp(-\phi_s)$. The other terms in (4.622) are the same as for the noninteracting electron gas.

By using the relation

$$\sinh^2(\lambda_p) = \frac{1}{2}[\cosh(2\lambda_p) - 1] = \frac{1}{2}\left[\frac{\omega_p}{E_p} - 1\right] \tag{4.623}$$

$$\phi_s(x) = \int_0^\infty \frac{dp}{p} e^{-\alpha p} \left(\frac{\omega_p}{E_p} - 1\right)[1 - \cos(px)] \tag{4.624}$$

which uses the prior result (4.609).

Any evaluation of the factor $\phi_s(x)$ must assume some specific form of the potential V_p between electrons. One possible model is to take $V_p = pV_0$, $V_0 =$ constant. This form of the potential is obtained from a delta function interaction in real space. This model assumes that the particles interact only when they directly collide. In this case ($\omega_p = pvF$)

$$E_p = p\sqrt{v_F^2 - V_0^2} \tag{4.625}$$

$$g \equiv \sinh^2(\lambda_p) = \frac{1}{2}\left[\frac{v_F}{\sqrt{v_F^2 - V_0^2}} - 1\right] \tag{4.626}$$

$$\phi_s(x) = 2g\int_0^\infty \frac{dp}{p} e^{-\alpha p}[1 - \cos(px)] \tag{4.627}$$

The factor $\sinh^2(\lambda_p)$ is a constant, which is called g. The integral for the exponential factor $\phi_s(x)$ is now simple to evaluate, since it has the same form as earlier for $\phi_0(x)$ in (4.582):

$$\phi_s(x) = g[\phi_0(x) + \phi_0^*(x)] = g\ln\left(1 + \frac{x^2}{\alpha^2}\right) \tag{4.628}$$

The delta function model makes the following prediction for the momentum distribution of the 1-particles:

$$\langle\Psi_1^\dagger(x)\Psi_1(x')\rangle = -\frac{e^{-ik_F(x-x')}}{2\pi i(x - x' + i\alpha)}\frac{1}{[1 + (x - x')^2/\alpha^2]^g} \tag{4.629}$$

$$\begin{aligned} n_{1,k} &= \int_{-\infty}^{\infty} dx e^{ik(x-x')}\langle\Psi_1^\dagger(x)\Psi_1(x')\rangle \\ &= -\int_{-\infty}^{\infty} \frac{dx}{2\pi i}\frac{e^{ix(k-k_F)}}{x + i\alpha}\frac{\alpha^{2g}}{(x^2 + \alpha^2)^g} \end{aligned} \tag{4.630}$$

Setting $g = 0$ recovers the noninteracting case $n_{1,k} = \Theta(k_F - k)$, which is obtained by closing the contour of integration in the UHP (upper half-plane) when $k_F < k$ and in the LHP (lower half-plane) when $k_F > k$. The pole at $x = -i\alpha$ is circled only in the latter case.

Mattis and Lieb (1965) showed that a more interesting result is found for the case where the coupling constant g is nonzero. Then $n_{1,k} =$ constant, independent of k, so the Fermi distribution is totally destroyed. This happens even in the limit where g is infinitesimally small. As $g \to 0$, then $n_{1,k} = \frac{1}{2}$.

This result is obtained by changing the integration variable to $y = x/\alpha$:

$$n_{1,k} = -\int_{-\infty}^{\infty} \frac{dy}{2\pi i}\frac{e^{iy\alpha(k-k_F)}}{y + i}\frac{1}{(y^2 + 1)^g} \tag{4.631}$$

The only α dependence is in the exponential factor. This exponential factor is needed for $g = 0$, since it tells us whether to close the integration contour in the upper or lower half-plane. However, for a nonzero value of g the integral converges even without the exponential factor. Therefore set $\alpha = 0$ before doing the integral and consider

$$n_{1,k} = -\int_{-\infty}^{\infty} \frac{dy}{2\pi i}\frac{1}{y + i}\frac{1}{(y^2 + 1)^g} \tag{4.632}$$

The right-hand side is no longer a function of $k - k_F$, and is a constant. The integral for $y > 0$ is added to that for $y < 0$ by changing the variable $y \to -y$ in the latter to give the real integral:

$$n_{1,k} = -\int_0^\infty \frac{dy}{2\pi i} \frac{1}{(y^2+1)^g} \left(\frac{1}{y+i} + \frac{1}{-y+i} \right) \tag{4.633}$$

$$= \int_0^\infty \frac{dy}{\pi} \frac{1}{(1+y^2)^{1+g}} = \frac{1}{2\sqrt{\pi}} \frac{\Gamma(\frac{1}{2}+g)}{\Gamma(1+g)} \tag{4.634}$$

The integral is in a standard form, which is given in tables [G&R, 3.194(3) after changing $y^2 = x$] in terms of gamma functions. In the limit where $g \to 0$ then $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$ so that

$$\lim_{g \to 0} n_{1,k} = \tfrac{1}{2} \tag{4.635}$$

The distribution function is a nonanalytic function of the coupling constant g. The usual noninteracting distribution function is found in the case of $g = 0$. The introduction of an arbitrarily weak delta function potential destroys the Fermi distribution, and each wave vector state is occupied with an equal probability. For the case where $g \to 0$, this probability approaches $\frac{1}{2}$. The $g = 0$ result is not obtained in the $g \to 0$ limit.

This result would be difficult to prove by perturbation theory and shows the value of an exact solution. These results pertain only to the one-dimensional electron gas. Behavior of this type is called *non-Fermi liquid behavior*. Fermi liquid theory is discussed in Chapter 11. Another name to describe such systems is *Luttinger liquids*.

4.6. POLARITONS

4.6.1. Semiclassical Discussion

The word *polariton* was coined by Hopfield (1958) to describe the normal modes in solids which propagate as electromagnetic waves. The word is a combination of *polarization* and *photon*, because these modes are combinations of free photons and the polarization modes of the solid. A new word was needed, because a new view was then emerging about the optical properties of solids. Hopfield popularized this new physics, although similar ideas had been discussed earlier by Fano (1956, 1960) and by Born and Huang (1954).

In the old view of electromagnetic wave propagation in solids, the light shone upon the surface of a sample and went into it. The polarization modes of the solid, e.g., TO phonons, could absorb some of this light.

The new view is that the light and the polarization modes in the solid are coupled into a new set of normal modes. These new modes are called polaritons. When light is shone upon the surface, polaritons are created which propagate inward. The mathematics is trivial; since both the photons and the polarization modes are usually described by harmonic oscillator equations, the new modes are obtained by solving coupled harmonic oscillator equations. The physical effect is semiclassical and need not involve quantum mechanics. The photon Green's function $\mathscr{D}_{\mu\nu}(\mathbf{q}, \omega)$ in a system with dielectric function $\varepsilon_{\mu\nu}(\mathbf{q}, \omega)$ was derived in (2.185). A

transverse wave will propagate with $\hat{\xi} \perp \mathbf{q}$, and the normal modes are the poles of the Green's functions. Assuming they exist, these poles are at

$$\omega^2 \varepsilon_t(\mathbf{q}, \omega) = \omega_{\mathbf{q}}^2 \equiv (qc)^2 \tag{4.636}$$

where $\varepsilon_t(\mathbf{q},\omega)$ is the transverse dielectric function.

The normal modes are the values of ω which satisfy this equation; call them Ω_q. For example, consider the phenomenological equation from optical phonons in ionic crystals for the transverse dielectric function, which is the same as the longitudinal dielectric function at long wavelength:

$$\varepsilon_t(\omega) = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{1 - \omega^2/\omega_{TO}^2} \tag{4.637}$$

When this form of the dielectric function is used in (4.636), there results a quadratic equation for ω^2. It may be solved in order to obtain the following two solutions:

$$\begin{aligned} \Omega_q^2 &= \tfrac{1}{2}(\tilde{\omega}_q^2 + \omega_{LO}^2) \pm \tfrac{1}{2}[(\tilde{\omega}_q^2 + \omega_{LO}^2)^2 - 4\tilde{\omega}_q^2\omega_{TO}^2]^{1/2} \\ \tilde{\omega}_q^2 &= \frac{(cq)^2}{\varepsilon_\infty} \end{aligned} \tag{4.638}$$

These solutions are plotted in Fig. 4.16. The solid lines are the solutions to (4.638) and are the actual normal modes in the solid. The dashed lines show the unperturbed photon mode $\tilde{\omega}_q$ and phonon mode ω_{TO}. It is these two which are coupled to form the new set of normal modes.

The picture of energy propagation is illustrated in Fig. 4.17. It shows a slab of polar material upon which light of frequency ω is incident I from the left. Some of the light is reflected R, while other parts may be transmitted through the slab and out the other side T. However, inside of the slab, the normal modes which propagate are those given by Eq. (4.638). The wave vectors of the polariton modes are determined by $\omega = \Omega_q$ and not by the free-photon value $q = (\omega/c)\sqrt{\varepsilon_\infty}$. This point is further illustrated by the observation that there are no polariton modes of real frequency in the range $\omega_{LO} > \omega > \omega_{TO}$, as shown in Fig. 4.16. Since there are no polariton modes in this gap, no energy can be transmitted through the slab (actually modes in the range $\omega_{LO} > \omega > \omega_{TO}$ have complex wave vector $q_R + iq_I = (\omega/c)\sqrt{\varepsilon(\omega)}$ so that energy can be transmitted through the slab of thickness d with the probability $\exp(-2dq_I)$, which is neglected). All the energy is reflected from the slab, as

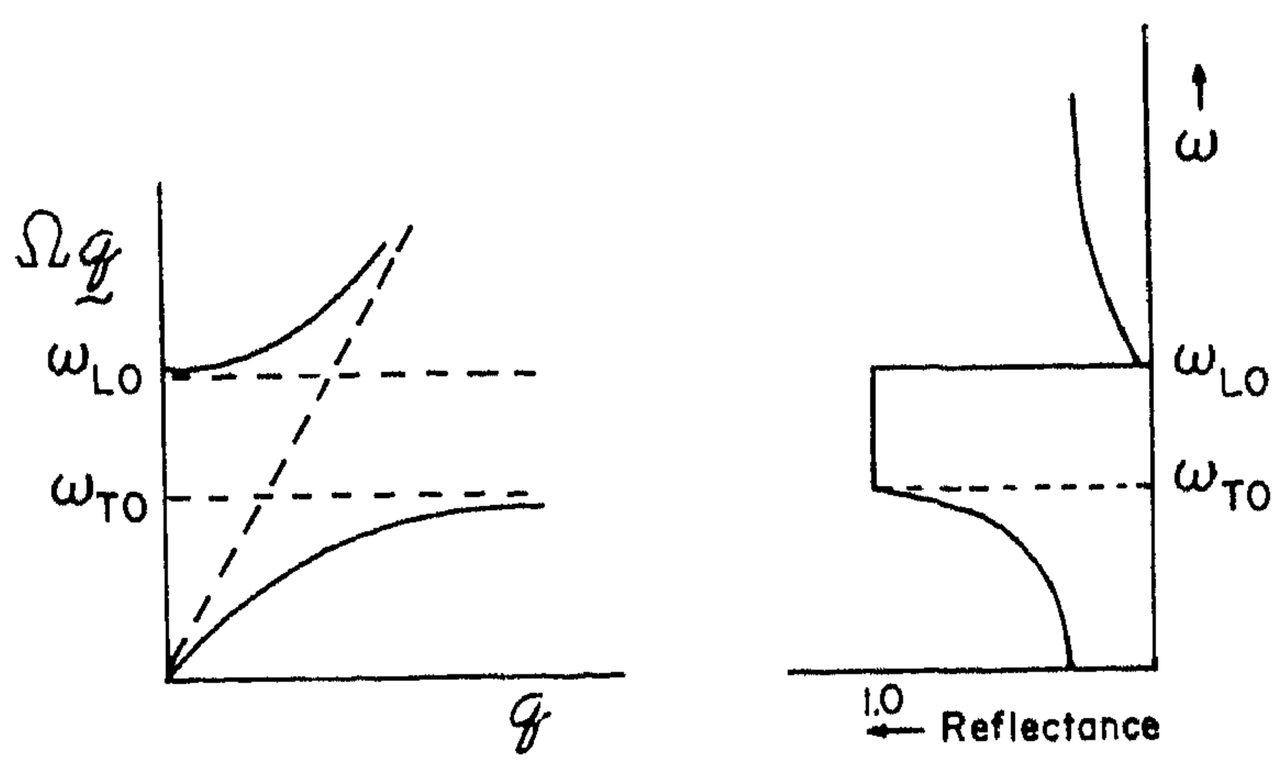


FIGURE 4.16 Polariton energy bands are shown at the left. There are no bands allowed for $\omega_{TO} < \omega < \omega_{LO}$, so that the solid must have a reflectivity of unity in this interval.

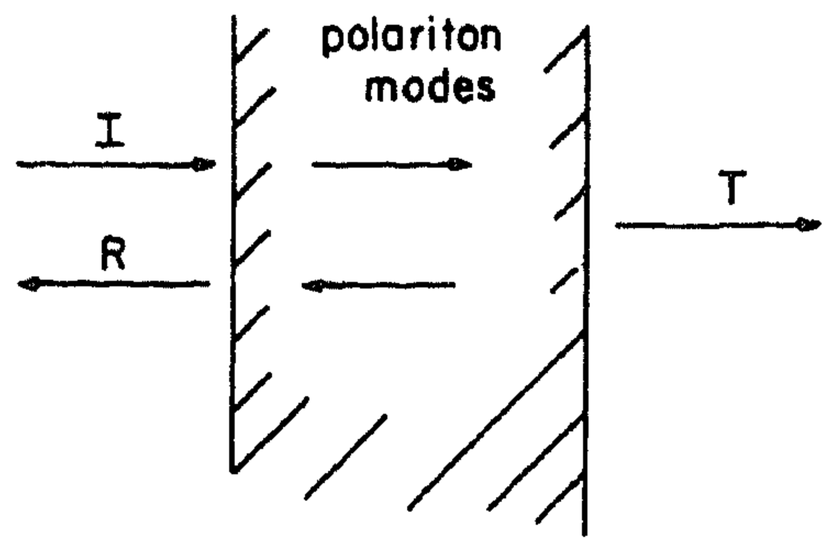


FIGURE 4.17 When light has normal incidence I upon a slab, there are the usual reflected R and transmitted components T. Inside the slab, the normal modes are polaritons.

shown in Fig. 4.16. The complete reflectivity for $\omega_{LO} > \omega > \omega_{TO}$ was known earlier for a simple oscillator in a solid, but the physics is clarified when cast into the form of an energy gap in the spectrum of normal modes.

The above equations have no damping in them, so that light is not absorbed in this model. All the electromagnetic energy, incident upon the sample, is either reflected or transmitted. The transmitted parts get carried through the slab by the polariton modes. The absorption of electromagnetic energy can be introduced into the equations by introducing damping for the phonons or additional scattering for the photons. However, if the TO-phonon system has no damping, then it does not cause energy absorption. The polariton picture is a rather different physical model than the older one in which the phonons caused the absorption of energy in the solid, and this absorption in turn led to reflection.

The most direct experimental verification of the theory is to measure the polariton dispersion curves, which has been done for the polaritons associated both with optical phonons and excitons. The first measurement was by Henry and Hopfield (1965), who measured the polaritons associated with the optical phonons in GaP. Their results are shown in Fig. 4.18, where the solid line is the theoretical curve (4.638), which compares well with their experimental points. They measured these modes by Raman scattering. The frequency of the Raman shift gave the frequency of the excited mode, while the directional change during the Raman scattering gives the information needed to deduce the wave vector of the polariton. There is obviously excellent agreement between theory and experiment.

4.6.2. Phonon–Photon Coupling

The semiclassical theory will be calculated using Green's functions. The self-energy operators for the photon, at least the parts due to the phonons, may be evaluated exactly, which gives the semiclassical theory. An interesting aspect to this derivation is that there are contributions from both the first- and second-order self-energy. Fortunately that is all, and the exact self-energy is found after the first two orders of perturbation theory.

The Hamiltonian has the form

$$H = H_0 + H' \tag{4.639}$$

$$H_0 = \omega_{TO} \sum_{\mathbf{q}} b_{\mathbf{q}}^{\dagger} b_{\mathbf{q}} + \sum_{\mathbf{q}\lambda} \omega_{\mathbf{q}} a_{\mathbf{q}\lambda}^{\dagger} a_{\mathbf{q}\lambda} \tag{4.640}$$

$$H' = -\frac{e}{Mc} \sum_j \mathbf{P}_j \cdot \mathbf{A}(\mathbf{R}_j) + \frac{e^2}{2Mc^2} \sum_j \mathbf{A}(\mathbf{R}_j) \cdot \mathbf{A}(\mathbf{R}_j) \tag{4.641}$$

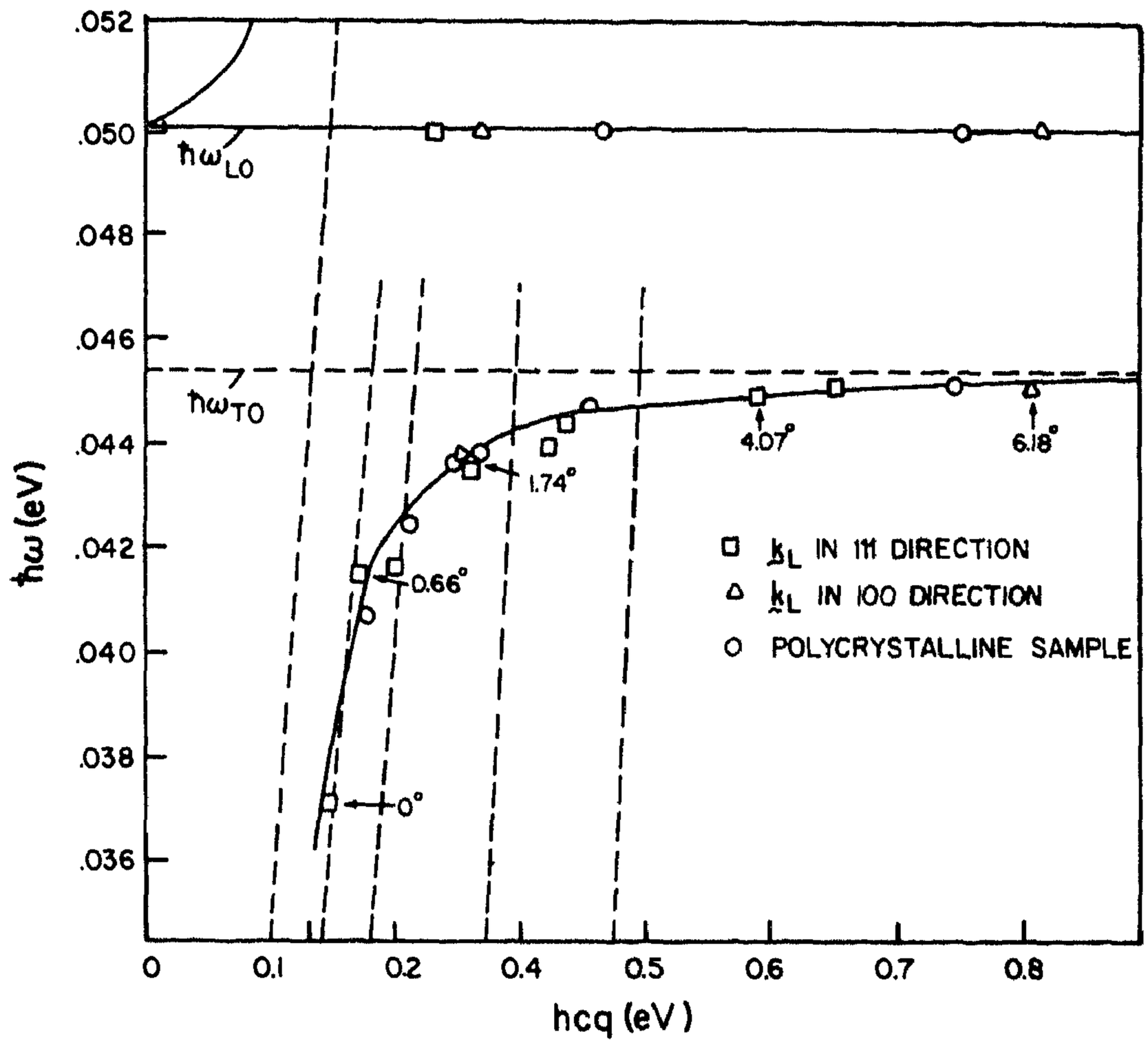


FIGURE 4.18 A plot of the observed energies and wave vectors of the polaritons and of the LO phonons in GaP. The theoretical dispersion curves are shown by the solid lines. The dispersion curves for the uncoupled dispersion curve are shown by the dashed lines. *Source*: Henry and Hopfield (1965) (used with permission).

The operators $b_{\mathbf{q}}, b_{\mathbf{q}}^{\dagger}$, are for phonons and $a_{\mathbf{q}\lambda}, a_{\mathbf{q}\lambda}^{\dagger}$ for photons. The interaction between the photons and optical phonons is contained in the term H'. The ions in the solid are treated as spherical balls with charge $+e$, equilibrium position $\mathbf{R}_j$, and momentum $\mathbf{P}_j$. The reduced mass M is used for the relative motions of positive and negative ions in the optical mode of oscillation. In terms of these coordinates, the current $\mathbf{J}_j$ and vector potential $\mathbf{A}(\mathbf{R}_j)$ are expressed (see Secs. 1.1 and 2.10) in terms of phonon and photon operators:

$$\mathbf{J}_j = \frac{e}{M}\mathbf{P}_j = ie\sum_{\mathbf{q}}\left(\frac{\omega_0}{2NM}\right)^{1/2}\hat{\xi}_{\mathbf{q}}(b_{\mathbf{q}}^{\dagger} - b_{-\mathbf{q}})e^{-i\mathbf{q}\cdot\mathbf{R}_j} = \frac{1}{\sqrt{N}}\sum_{\mathbf{q}}\mathbf{J}(\mathbf{q})e^{-i\mathbf{q}\cdot\mathbf{R}_j}$$
$$J_{\mu}(\mathbf{q}) = ie\left(\frac{\omega_0}{2M}\right)^{1/2}\hat{\xi}_{\mathbf{q}}(b_{\mathbf{q}}^{\dagger} - b_{-\mathbf{q}}) \tag{4.642}$$

$$A_{\mu}(\mathbf{R}_j) = \sum_{\mathbf{q}\lambda}\left(\frac{2\pi c^2}{\omega_{\mathbf{q}}\mathcal{V}}\right)^{1/2}\xi_{\mu}e^{i\mathbf{q}\cdot\mathbf{R}_j}(a_{\mathbf{q}\lambda} + a_{-\mathbf{q}\lambda}^{\dagger}) = \frac{c}{\sqrt{\mathcal{V}}}\sum_{\mathbf{q}}A_{\mu}(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{R}_j}$$
$$A_{\mu}(\mathbf{q}) = \sum_{\lambda}\left(\frac{2\pi}{\omega_{\mathbf{q}}}\right)^{1/2}\xi_{\mu\lambda}(a_{\mathbf{q}\lambda} + a_{-\mathbf{q}\lambda}^{\dagger}) \tag{4.643}$$

The first self-energy is from the photon self-energy, which arises from the term in A^2. The interaction in the second term in (4.641) is

$$H_{A^2} = \frac{e^2}{2M\mathcal{V}_0}\sum_{\mathbf{q},\mu}A_{\mu}(\mathbf{q})A_{\mu}(-\mathbf{q}) \tag{4.644}$$

where $v_0 = v/N$ is the volume of the unit cell in a binary crystal of positive and negative ions; i.e., v_0 is the volume per phonon mode. The self-energy comes from the first-order term in the expansion of the S matrix:

$$\begin{aligned}\mathscr{D}_{\mu\nu}(\mathbf{k}, i\omega) &= -\int_0^\beta d\tau e^{i\omega\tau}\langle T_\tau A_\mu(\mathbf{k}, \tau)A_\nu(-\mathbf{k}, 0)\rangle \\ &= \mathscr{D}^{(0)}_{\mu\nu}(\mathbf{k}, i\omega) + \frac{e^2}{2Mv_0}\int_0^\beta d\tau e^{i\omega\tau}\int_0^\beta d\tau_1 \sum_{\mathbf{q},\delta} \\ &\quad \times \langle T_\tau \hat{A}_\mu(\mathbf{k}, \tau)\hat{A}_\delta(\mathbf{q}, \tau_1)\hat{A}_\delta(-\mathbf{q}, \tau_1)\hat{A}_\nu(-\mathbf{k}, 0)\rangle + \cdots \\ A_\mu(\mathbf{k}, \tau) &= e^{\tau H}A_\mu e^{-\tau H} \\ \hat{A}_\mu(\mathbf{k}, \tau) &= e^{\tau H_0}A_\mu e^{-\tau H_0}\end{aligned} \tag{4.645}$$

The four-operator sequence in the interaction term is equal to

$$\begin{aligned}&\int_0^\beta d\tau e^{i\omega\tau}\int_0^\beta d\tau_1 \sum_{\mathbf{q},\delta}\langle T_\tau \hat{A}_\mu(\mathbf{k}, \tau)\hat{A}_\delta(\mathbf{q}, \tau_1)\hat{A}_\delta(-\mathbf{q}, \tau_1)\hat{A}_\nu(-\mathbf{k}, 0)\rangle \\ &\quad = 2\sum_\delta \mathscr{D}^{(0)}_{\mu\delta}(\mathbf{k}, i\omega)\mathscr{D}^{(0)}_{\delta\nu}(\mathbf{k}, i\omega)\end{aligned} \tag{4.646}$$

The factor of 2 arises because of the two choices of pairing: $\mathbf{k} = \mathbf{q}$ or $\mathbf{k} = -\mathbf{q}$. The series in (4.645) is

$$\begin{aligned}\mathscr{D}_{\mu\nu}(\mathbf{k}, i\omega) &= \mathscr{D}^{(0)}_{\mu\nu}(\mathbf{k}, i\omega) + \Pi^{(1)}\sum_\delta \mathscr{D}^{(0)}_{\mu\delta}(\mathbf{k}, i\omega)\mathscr{D}^{(0)}_{\delta\nu}(\mathbf{k}, i\omega) + \cdots \\ \Pi^{(1)} &= \frac{e^2}{Mv_0}\end{aligned} \tag{4.647}$$

According to (2.184) the self-energy $\Pi^{(1)}(\mathbf{k}, i\omega)$ enters into the denominator of the photon Green's function with an additional factor of 4π:

$$\mathscr{D}_{\mu\nu}(\mathbf{k}, i\omega) = \frac{4\pi(\delta_{\mu\nu} - k_\mu k_\nu/k^2)}{(i\omega)^2 - \omega_{\mathbf{k}}^2 - 4\pi\Pi^{(1)}} \tag{4.648}$$

$$4\pi\Pi^{(1)} = \frac{4\pi e^2}{Mv_0} = \omega_p^2 \tag{4.649}$$

The quantity $4\pi\Pi^{(1)} = \omega_p^2$ is just a plasma frequency. It would be the longitudinal frequency $\omega_L = \omega_p$ of the ion vibrational oscillation if there were no other restoring force. Of course, there is a restoring force, which leads to the transverse frequency $\omega_{TO} \equiv \omega_0$. The longitudinal frequency is then determined by the combination of the short-range restoring force ω_{TO}^2 and the long-range Coulomb force ω_p^2, $\omega_{LO}^2 = \omega_{TO}^2 + \omega_p^2$. All this discussion is slightly irrelevant because we are solving for transverse modes, not longitudinal modes. For the transverse case, the modes are at the poles of the Green's function in (4.648). The poles are at the frequencies $\omega^2 = \omega_k^2 + 4\pi\Pi_{\text{ret}}(\mathbf{k}, \omega)$, which would be at $c^2k^2 + \omega_p^2$ if there were no other contributions to the self-energy. However, there is another contribution to $\Pi(\mathbf{k}, \omega)$ from the $\mathbf{P}\cdot\mathbf{A}$ term in H', which cancels ω_p^2.

The A^2 term in the interaction H' does not contribute any further terms to the self-energy of the photon. The self-energy contribution $\Pi^{(1)}$ provides an exact evaluation of this term in the Hamiltonian. This exact result could have been deduced without the aid of Green's

functions. The Hamiltonian for free photons can be written in terms of harmonic oscillator coordinates by identifying the displacement $q_\mu(\mathbf{k}) = A_\mu(\mathbf{k})/\sqrt{4\pi}$. The Hamiltonian has the form

$$H_{0,\text{photon}} = \frac{1}{2}\sum_{\mathbf{k}}[\mathbf{p}(\mathbf{k})\cdot\mathbf{p}(-\mathbf{k}) + \omega_k^2\mathbf{q}(\mathbf{k})\cdot\mathbf{q}(-\mathbf{k})] \tag{4.650}$$

When the A^2 term is written in the same notation, it is $\omega_p^2\mathbf{q}(\mathbf{k})\cdot\mathbf{q}(-\mathbf{k})/2$. Add the A^2 term to H_0, and then the new potential energy is $\Omega_k^2\mathbf{q}(\mathbf{k})\cdot\mathbf{q}(-\mathbf{k})/2$, where the new modes are $\Omega_k^2 = \omega_k^2 + \omega_p^2$. The same result was obtained using Green's functions.

The next step is to evaluate the self-energy arising from the $\mathbf{J}\cdot\mathbf{A}$ term in the interaction H' in (4.641). The symbol $\mathbf{J}_j$ denotes the current of the vibrating ion pair. The $\mathbf{J}\cdot\mathbf{A}$ term is rewritten in wave vector space as

$$H_{p\cdot A} = -\frac{1}{\nu_0}\sum_{\mathbf{k}}\mathbf{J}(\mathbf{k})\cdot\mathbf{A}(-\mathbf{k}) \tag{4.651}$$

One factor in the self-energy expressions for the photon is the correlation function of the ion momentum $J_\mu(\mathbf{k})$ with itself. Define a correlation function

$$\begin{aligned}\Phi_{\mu\nu}(\mathbf{k}, i\omega) &= -\int_0^\beta d\tau e^{i\omega\tau}\langle T_\tau[e^{\tau H_0}J_\mu(\mathbf{k})e^{-\tau H_0}]J_\nu(-\mathbf{k})\rangle \\ &= \delta_{\mu\nu}\frac{e^2\omega_0}{2M}\int_0^\beta d\tau e^{i\omega\tau}[e^{\tau\omega_0}\langle b_{\mathbf{k}}^\dagger b_{\mathbf{k}}\rangle + e^{-\tau\omega_0}\langle b_{\mathbf{k}}b_{\mathbf{k}}^\dagger\rangle] \\ &= \delta_{\mu\nu}\frac{e^2}{M}\frac{\omega_0^2}{(i\omega_n)^2 - \omega_0^2}\end{aligned} \tag{4.652}$$

which is easily evaluated since the properties of the phonon operators are known.

The $\mathbf{J}\cdot\mathbf{A}$ interaction does not contribute any term to the first-order self-energy expression $\Pi^{(1)}$, because the first-order term has only a single factor of the operator $J_\mu(\mathbf{k})$, and the average of this operator is zero. The second-order term in the expansion for the S matrix is

$$\begin{aligned}&\frac{1}{2!\nu_0}\int_0^\beta d\tau e^{i\omega\tau}\int_0^\beta d\tau_1\int_0^\beta d\tau_2\sum_{\mathbf{k}_1\mathbf{k}_2}\sum_{\delta\delta'}\langle T_\tau\hat{J}_\delta(\mathbf{k}_1,\tau_1)\hat{J}_{\delta'}(\mathbf{k}_2,\tau_2)\rangle \\ &\quad\times\langle T_\tau\hat{A}_\mu(\mathbf{k},\tau)\hat{A}_\delta(\mathbf{k}_1,\tau_1)\hat{A}_{\delta'}(\mathbf{k}_2,\tau_2)\hat{A}_\nu(-\mathbf{k},0)\rangle\end{aligned}$$

which is evaluated by the usual rules to give

$$\frac{1}{\nu_0}\sum_{\delta\delta'}\mathscr{D}_{\mu\delta}^{(0)}(\mathbf{k}, i\omega)\Phi_{\delta\delta'}(\mathbf{k}, i\omega)\mathscr{D}^{(0)}(\mathbf{k}, i\omega) \tag{4.653}$$

where $\Phi_{\delta\delta'}(\mathbf{k}, i\omega)$ is given in (4.652). A factor of 2 comes from the two possible pairings $\mathbf{k} = \mathbf{k}_1$ or $\mathbf{k} = \mathbf{k}_2$. This self-energy expression is

$$\Pi^{(2)}(\mathbf{k}, i\omega) = \frac{1}{\nu_0}\Phi_{\mu\nu}(\mathbf{k}, i\omega) = \frac{\delta_{\mu\nu}}{4\pi}\frac{\omega_p^2\omega_0^2}{(i\omega_n)^2 - \omega_0^2} \tag{4.654}$$

This self-energy term is the only one from this interaction. The total self-energy is obtained by adding the two contributions:

$$4\pi\Pi(\mathbf{k}, i\omega) = 4\pi[\Pi^{(1)} + \Pi^{(2)}] = \omega_p^2\left[1 + \frac{\omega_0^2}{(i\omega)^2 - \omega_0^2}\right] = \frac{(i\omega)^2\omega_p^2}{(i\omega)^2 - \omega_0^2}$$

The photon self-energy for the Hamiltonian (4.641) has been found exactly. The energy denominator in the photon Green's function (4.648) can now be written in terms of the dielectric function:

$$\begin{aligned}(i\omega)^2 - \omega_k^2 - 4\pi\Pi(\mathbf{k}, i\omega) &= (i\omega)^2\varepsilon(i\omega) - \omega_k^2 \\ \varepsilon(i\omega) &= 1 + \frac{\omega_p^2}{\omega_0^2 - (i\omega)^2} \\ \mathscr{D}_{\mu\nu}(\mathbf{k}, i\omega) &= \frac{4\pi(\delta_{\mu\nu} - k_\mu k_\nu/k^2)}{(i\omega)^2\varepsilon(i\omega) - \omega_k^2}\end{aligned} \tag{4.655}$$

This dielectric function should be compared with (4.637) from optical phonons. The present discussion assumes that $\varepsilon_\infty = 1$. Otherwise, a value of ε_∞ other than unity will modify the vector potential with factors of ε_∞. Similarly, identify the static dielectric constant as $\varepsilon_0 = 1 + \omega_p^2/\omega_0^2$. Another feature of the dielectric function (4.655) is that the frequency of the longitudinal mode oscillation is determined by the condition that $\varepsilon(\omega_{LO}) = 0$, which has the solution $\omega_{LO}^2 = \omega_0^2 + \omega_p^2$. The transverse frequency is $\omega_{TO} = \omega_0$. With these identifications, the two expressions (4.637) and (4.655) for the dielectric functions are identical.

The Green's function solution of the Hamiltonian for the coupled phonon–photon modes yields exactly the same equation $\omega^2\varepsilon(\omega) = \omega_k^2$ as found from the semiclassical arguments in Sec. 4.6.1. The two modes, photon and phonon, are coupled and form two new modes which are polaritons. In fact, the original Hamiltonian (4.641) could be solved by a canonical transformation to a new set of operators, say α_k and β_k, which diagonalize the Hamiltonian. These new creation and destruction operators are for the polariton modes. This exercise is assigned as a problem.

4.6.3. Exciton–Photon Coupling

The polariton concept applies to the mixing of any polarization waves with the photons. Although the phonons are an obvious example, the physics is the same for waves which are associated with electronic polarization. Here we analyze a simple case of this phenomena. The solid is treated as a collection of atoms which are only weakly interacting, say by van der Waals forces. The electrons in each atom can occupy a set of discrete energy levels ε_m. This model is applied to molecular solids, in which molecules are weakly bound together, and the optical properties of the solid are often only weak perturbations on the optical properties of the free molecules (see Davydov, 1971, or Mahan, 1975).

A light wave, which is traveling through the solid, can induce the electrons to change their electronic state from n to m. This excitation process can really happen if the photon energy $\hbar\omega$ matches the energy of electronic excitation $\hbar\omega = \varepsilon_m - \varepsilon_n$. However, for photon frequencies ω which do not match any of the excitation frequencies $\omega \neq \omega_\alpha = \varepsilon_m - \varepsilon_n$, real excitations cannot occur. Then the light wave will induce a virtual transition to this excited state: a virtual process is an excitation which starts to occur but violates energy conservation,

so the excitation process ceases. Nevertheless, the electronic system is polarized during this virtual excitation process. The atom will have electric dipole moment μ whose magnitude is proportional to the amplitude of the electric field of the light which is acting upon the atom:

$$\mu_\nu = \sum_\lambda \alpha_{\nu\lambda}(\omega) E_\lambda e^{-i\omega t} \tag{4.656}$$

An expression for the atomic polarizability $\alpha_{\nu\lambda}(\omega)$ will be derived below. The Hamiltonian for each atom of atomic number Z will be taken to have the nonrelativistic form

$$H = \sum_i \frac{p_i^2}{2m} - Ze^2 \sum_i \frac{1}{r_i} + \frac{e^2}{2} \sum_{\langle ij \rangle} \frac{1}{r_{ij}} \tag{4.657}$$

It is assumed to have solutions which are single-particle orbitals $\phi_n(\mathbf{r})$ with energy ε_n. In the Hartree–Fock approximation, the state of the atom is a Slater determinant of the occupied orbital states. The ground state is the Slater determinant of the set of orbitals with the lowest energy. An excited electronic state is obtained by raising an electron from one of the ground state orbitals to another one of higher energy, which is unoccupied in the ground state. Denote by Greek subscripts α or β the possible excitation energies $\omega_\alpha = \varepsilon_m - \varepsilon_n$. Each excitation α corresponds to moving one electron from a ground state to an excited state orbital. The operator $b_\alpha^\dagger$ is introduced to describe this excitation, which can be defined in terms of the operators a_n of the electronic states:

$$b_\alpha^\dagger = a_m^\dagger a_n \tag{4.658}$$

The a_n operators are fermions. The b_α operators are the product of two fermion operators and have boson properties. However, they are not bosons since the number in each state is limited: one has that $(b_n^\dagger)^N = 0$, where the value N depends on the maximum number of electrons which are in state n or may be in state m. However, the b_α operators are approximated as pure bosons. It is an adequate approximation as long as the density of such bosons is low, so that the average number of bosons on each atomic site is much less than unity. Then the restriction on the site occupancy of the bosons is irrelevant, and the b_α operators obey good boson statistics. The Hamiltonian of a single atom is approximated as

$$H = E_g + \sum_\alpha \hbar\omega_\alpha b_\alpha^\dagger b_\alpha \tag{4.659}$$

$$[b_\alpha, b_\beta^\dagger] = \delta_{\alpha\beta} \tag{4.660}$$

where E_g is the ground state energy and $b_\alpha^\dagger b_\alpha$ is the number of excitations.

The model of an atom is far too simple by today's theoretical standards. The Hartree–Fock (HF) approximation is seldom adequate to describe an atom, and usually one has to take a linear combination of HF states (configurational interaction, or CI in the chemist's notation). The single-particle orbitals $\phi_n(\mathbf{r})$ also vary with the excited state configuration, which requires a more complicated Hamiltonian than (4.659) in order to describe excitations. These complications are omitted, which hopefully are only a side issue for the discussion of excitons.

An important quantity for our discussion is the dipole matrix element $\mathbf{X}(\alpha)$ for the states ϕ_n and ϕ_m which comprise the transition

$$\mathbf{X}(\alpha) = \int d^3r \phi_m^*(\mathbf{r}) \mathbf{x} \phi_n(\mathbf{r}) \tag{4.661}$$

This matrix element will be zero in most cases, since it is nonzero only when the two electronic orbitals ϕ_n and ϕ_m differ by one unit of angular momentum. This matrix element is important for the case the excitations occur by the interaction with light. The rate of excitation is determined by the dipole matrix element.

Now introduce an operator P_μ which is the polarization operator for the atom, where μ denotes vector direction and α denotes excitation:

$$P_\mu = e\sum_\mu X_\mu(\alpha)(b_\alpha + b_\alpha^\dagger) \tag{4.662}$$

It should not be confused with the momentum, which has the same symbol. This operator has the units of polarization, which is charge times length. One can think of the quantity $X_\mu(b_\alpha + b_\alpha^\dagger)$ as a typical displacement operator, which in this case refers to the displacement of electronic charge in the atom. It is quite analogous to the displacement operator for the harmonic oscillator system, which is $X_{\mathbf{q}}(a_{\mathbf{q}} + a_{-\mathbf{q}}^\dagger)$, where $X_{\mathbf{q}} = (\hbar/2M\omega_{\mathbf{q}})^{1/2}$ is the unit of length.

The symbol $\mathscr{F}_{\mu\nu}(i\omega)$ is used to denote the correlation function of the polarization operator P_μ with itself:

$$\mathscr{F}_{\mu\nu}(i\omega) = -\int_0^\beta d\tau e^{i\omega\tau}\langle T_\tau P_\mu(\tau)P_\nu(0)\rangle \tag{4.663}$$

It is easily evaluated for the noninteracting Hamiltonian (4.659):

$$\mathscr{F}_{\mu\nu}^{(0)}(i\omega) = -e^2\int_0^\beta d\tau e^{i\omega\tau}\sum_\alpha X_\mu(\alpha)X_\nu(\alpha)(e^{-\tau\omega_\alpha}\langle b_\alpha b_\alpha^\dagger\rangle + e^{\tau\omega_\alpha}\langle b_\alpha^\dagger b_\alpha\rangle) \tag{4.664}$$

$$\mathscr{F}_{\mu\nu}^{(0)}(i\omega) = \sum_\alpha \frac{2e^2X_\mu X_\nu\omega_\alpha}{(i\omega_n)^2 - \omega_\alpha^2} \tag{4.665}$$

The quantity $\mathscr{F}_{\mu\nu}^{(0)}(i\omega)$ is actually the polarizability $\alpha_{\mu\nu}(i\omega)$ of the atom. To keep this discussion simple, assume that this quantity is isotropic:

$$\mathscr{F}_{\mu\nu}^{(0)}(i\omega) = \alpha_{\mu\nu}(i\omega) = \delta_{\mu\nu}\alpha(\omega) \tag{4.666}$$

$$\alpha(i\omega) = 2e^2\sum_\alpha \frac{X_\mu^2\omega_\alpha}{\omega_\alpha^2 - (i\omega_n)^2} \tag{4.667}$$

So far the discussion has been confined to the properties of a single atom. Next consider the additional properties which result when the atoms are collected together in the solid. The discussion about photons is deferred to later, and now consider only the original Hamiltonian (4.657) as applied to a collection of atoms. The electrons on one atom will interact with the nucleus and electrons on adjacent atoms. Denote the position of an electron $\mathbf{r}_i = \mathbf{R}_i + \mathbf{x}_i$, where $\mathbf{R}_i$ is the position of the atom and $\mathbf{x}_i$ is the displacement of the electron from the center

of the atom. For two electrons on different atoms, the interaction is expanded assuming that $|\mathbf{R}_i - \mathbf{R}_j| > |\mathbf{x}_i - \mathbf{x}_j|$:

$$\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} = \frac{1}{|\mathbf{R}_i - \mathbf{R}_j|} + \frac{(\mathbf{x}_i - \mathbf{x}_j)\cdot(\mathbf{R}_i - \mathbf{R}_j)}{|\mathbf{R}_i - \mathbf{R}_j|^3} + \tfrac{1}{2}\sum(\mathbf{x}_i - \mathbf{x}_j)_\mu \phi_{\mu\nu}(\mathbf{R}_i - \mathbf{R}_j)(\mathbf{x}_i - \mathbf{x}_j)_\nu + \cdots \tag{4.668}$$

$$\phi_{\mu\nu}(\mathbf{R}) = \frac{\delta_{\mu\nu}}{R^3} - \frac{3R_\mu R_\nu}{R^5}$$

The first two terms can usually be neglected, since the first is a constant, and the second will vanish for crystals with inversion symmetry. The first important term, resulting from the interaction between electrons on different atoms, is the dipole–dipole interaction:

$$x_{i\mu}\phi_{\mu\nu}(\mathbf{R}_i - \mathbf{R}_j)x_{j\nu} \tag{4.669}$$

The standard practice is to retain only this term in the theory, which is called the dipole approximation. Sometimes it is improved by also including the next terms, which involve quadrupoles and octopoles. For a treatment of exciton theory without making a monopole expansion but instead retaining all the terms in the electron–electron interaction, see Agranovitch (1960, 1961) for the formalism and Mahan (1975) for a sample calculation.

The notation needs to be expanded to include the site l of the atom in the solid. The exciton operators are referred to as $b_{\alpha l}$ and polarization operators as $P_\mu(l)$. For one atom per unit cell, each quantity may also be expanded in wave vector space:

$$P_\mu(l) = e\sum_\alpha X_\mu(\alpha)(b_{\alpha,l} + b^\dagger_{\alpha,l}) \tag{4.670}$$

$$P_\mu(\mathbf{k}) = \frac{1}{\sqrt{N}}\sum_l e^{i\mathbf{k}\cdot\mathbf{R}_l}P_\mu(l) = e\sum_\alpha X_\mu(\alpha)(b_{\alpha,\mathbf{k}} + b^\dagger_{\alpha,-\mathbf{k}}) \tag{4.671}$$

$$b_{\alpha,\mathbf{k}} = \frac{1}{\sqrt{N}}\sum_l e^{i\mathbf{k}\cdot\mathbf{R}_l}b_{\alpha,l} \tag{4.672}$$

In the dipole–dipole interaction (4.669), the dipole moment of the electron $ex_{i\mu}$ is replaced by the equivalent displacement operator $P_\mu(i)$. The dipole–dipole interaction between excitations on different atoms becomes:

$$\frac{e^2}{2}\sum_{ij\mu\nu} x_{i\mu}\phi_{\mu\nu}(\mathbf{R}_i - \mathbf{R}_j)x_{j\nu} \to \frac{1}{2}\sum_{ij\mu\nu} P_\mu(i)\phi_{\mu\nu}(\mathbf{R}_i - \mathbf{R}_j)P_\nu(j)$$

$$= \frac{2\pi}{v_0}\sum_{\mu\nu\mathbf{k}} P_\mu(\mathbf{k})T_{\mu\nu}(\mathbf{k})P_\nu(-\mathbf{k}) \tag{4.673}$$

$$T_{\mu\nu}(\mathbf{k}) = \frac{v_0}{4\pi}\sum_{j\neq 0} e^{i\mathbf{k}\cdot\mathbf{R}_j}\phi_{\mu\nu}(\mathbf{R}_j) \tag{4.674}$$

The interaction between excitations on the same atom is omitted, since this contribution is already included implicitly in the free-atom Hamiltonian (4.657). The last step above gives the interaction in wave vector notation. The excitations in wave vector space interact with the

lattice transform of the dipole–dipole interaction, which is called $T_{\mu\nu}(\mathbf{k})$. This function is only needed at small wave vectors, which for cubic crystals is (Cohen and Keffer, 1955)

$$\lim_{\mathbf{k}\to 0} T_{\mu\nu}(\mathbf{k}) = \frac{k_\mu k_\nu}{k^2} - \tfrac{1}{3}\delta_{\mu\nu} + O(ka) \tag{4.675}$$

where a is a lattice constant. If one tries to evaluate this function exactly at the point $\mathbf{k} = 0$, it amounts to summing the dipole–dipole interaction over neighboring sites, which gives zero. The function is conditionally convergent, since the limit as $\mathbf{k}$ goes to zero gives a different result than at the point that $\mathbf{k} = 0$. The limiting value is correct.

The Hamiltonian under discussion is the original atomic part (4.659) plus the dipole–dipole interaction term. The atomic part can also be expressed in wave vector notation, so consider

$$\begin{aligned} H &= \sum_{\alpha\mathbf{k}} \omega_\alpha b^\dagger_{\alpha,\mathbf{k}} b_{\alpha,\mathbf{k}} + \frac{1}{2}\sum_{\alpha\beta,\mathbf{k}} V_{\alpha\beta}(\mathbf{k})(b_{\alpha,\mathbf{k}} + b^\dagger_{\alpha,-\mathbf{k}})(b_{\beta,-\mathbf{k}} + b^\dagger_{\beta,\mathbf{k}}) \\ V_{\alpha\beta}(\mathbf{k}) &= \frac{4\pi e^2}{v_0}\sum_{\mu\nu} X_\mu(\alpha) T_{\mu\nu}(\mathbf{k}) X_\nu(\beta) \end{aligned} \tag{4.676}$$

This form of a Hamiltonian has been diagonalized before, see (4.438). However, the present case is more complicated, because of the summation over the different excitations α and β. Equation (4.676) is solved by the use of a Green's function. The correlation function (4.663) is redefined to include wave vectors:

$$\mathscr{F}_{\mu\nu}(\mathbf{k}, i\omega) = -\int_0^\beta d\tau e^{i\omega\tau} \langle T_\tau P_\mu(\mathbf{k}, \tau) P_\nu(-\mathbf{k}, 0)\rangle \tag{4.677}$$

This correlation function is evaluated by a diagrammatic expansion. The first term in the Hamiltonian (4.676) is treated as H_0, and the second term is the interaction V. For this H_0 the unperturbed Green's function was evaluated earlier in (4.664) and is the negative of the atomic polarizability:

$$\mathscr{F}^{(0)}_{\mu\nu}(\mathbf{k}, i\omega) = \sum_\alpha \frac{2e^2 X_\mu X_\nu \omega_\alpha}{(i\omega_n)^2 - \omega_\alpha^2} = -\alpha_{\mu\nu}(i\omega_n) \tag{4.678}$$

The self-energy term for the perturbation V occurs in first order. An important feature of the evaluation of this term is the separability of the potential $V_{\alpha\beta}$,

$$V = \frac{4\pi}{v_0}\sum_{\mu\nu\mathbf{k}} P_\mu(\mathbf{k}) T_{\mu\nu}(\mathbf{k}) P_\nu(-\mathbf{k}) \tag{4.679}$$

so that the different α and β summations just go into the factors $P_\mu(\mathbf{k})$ and $P_\nu(-\mathbf{k})$. For example, the first-order term in the expansion of the S matrix is

$$\begin{aligned} \mathscr{F}_{\mu\nu}(\mathbf{k}, i\omega) &= \mathscr{F}^{(0)}_{\mu\nu} \int_0^\beta d\tau e^{i\omega\tau} \int_0^\beta d\tau_1 \sum_{\mu'\nu'\mathbf{k}'} \frac{2\pi}{v_0} T_{\mu'\nu'}(\mathbf{k}') \\ &\quad \times \langle T_\tau \hat{P}_\mu(\mathbf{k}, \tau) \hat{P}_{\mu'}(-\mathbf{k}', \tau_1) \hat{P}_{\nu'}(\mathbf{k}', \tau_1) \hat{P}_\nu(-\mathbf{k}, 0)\rangle + \cdots \\ &= \mathscr{F}^{(0)}_{\mu\nu}(\mathbf{k}, i\omega) + \frac{4\pi}{v_0}\sum_{\mu'\nu'} \mathscr{F}^{(0)}_{\mu\mu'}(\mathbf{k}, i\omega) T_{\mu'\nu'}(\mathbf{k}) \mathscr{F}^{(0)}_{\nu'\nu}(\mathbf{k}, i\omega) + \cdots \end{aligned}$$

The self-energy is $4\pi T_{\mu\nu}/v_0$. It is the only self-energy term, and higher-order terms in the S-matrix expansion just produce multiples of this term. The Dyson equation for the correlation function is the matrix equation

$$\mathscr{F}_{\mu\nu}(\mathbf{k}, i\omega) = \mathscr{F}^{(0)}_{\mu\nu}(\mathbf{k}, i\omega) + \frac{4\pi}{v_0}\sum_{\mu'\nu'} \mathscr{F}^{(0)}_{\mu\mu'}(\mathbf{k}, i\omega) T_{\mu'\nu'}(\mathbf{k}) \mathscr{F}_{\nu'\nu}(\mathbf{k}, i\omega) \tag{4.680}$$

This equation is not too hard to solve in the general case, since it is only a 3×3 matrix in the direction variables $(\mu, \nu) = (x, y, z)$. We shall only consider the simplest possible case, which has an isotropic polarizability $\mathscr{F}^{(0)}_{\mu\nu} = -\delta_{\mu\nu}\alpha$ and the long-wavelength form of the interaction $T_{\mu\nu}$ as given in (4.675). Then the equation to be solved simplifies to

$$\mathscr{F}_{\mu\nu} = -\delta_{\mu\nu}\alpha - \frac{4\pi\alpha}{v_0}\sum_{\lambda}\left(\frac{k_\mu k_\lambda}{k^2} - \frac{\delta_{\mu\lambda}}{3}\right)\mathscr{F}_{\lambda\nu} \tag{4.681}$$

At small values of wave vector the matrix $\mathscr{F}_{\mu\nu}$ depends only on the scalar functions C and D,

$$\mathscr{F}_{\mu\nu} = C\delta_{\mu\nu} + D\frac{k_\mu k_\nu}{k^2} \tag{4.682}$$

We can generate equations in only C and D by inserting the above results in (4.681) and then equating coefficients of the terms $\delta_{\mu\nu}$ and $k_\mu k_\nu/k^2$

$$C = -\alpha + \frac{4\pi\alpha}{3v_0}C \tag{4.683}$$

$$D = -\frac{4\pi\alpha}{v_0}(C + \tfrac{2}{3}D) \tag{4.684}$$

These two equations for C and D can be solved algebraically:

$$\begin{aligned} C &= -\frac{\alpha}{1 - 4\pi\alpha/3v_0} \\ D &= \frac{\alpha(4\pi\alpha/v_0)}{(1 - 4\pi\alpha/3v_0)(1 + 8\pi\alpha/3v_0)} \\ \mathscr{F}_{\mu\nu} &= -\frac{\alpha}{1 - 4\pi\alpha/3v_0}\left[\delta_{\mu\nu} - \frac{4\pi\alpha}{1 + 8\pi\alpha/3v_0}\frac{k_\mu k_\nu}{k^2}\right] \end{aligned} \tag{4.685}$$

Equation (4.685) is the long-wavelength form of $\mathscr{F}_{\mu\nu}(\mathbf{k}, i\omega)$.

The quantity $-\mathscr{F}_{\mu\nu}$ is the polarizability of the solid, just as $-\mathscr{F}^{(0)}_{\mu\nu}$ is the polarizability of the atom. The interactions in V modify the atomic polarizability. These modifications are just the local field corrections. In a cubic dielectric, the dielectric function at long wavelength has the Lorenz–Lorentz form:

$$\varepsilon(\omega) = 1 + \frac{4\pi\alpha/v_0}{1 - 4\pi\alpha/3v_0} \tag{4.686}$$

The factor $4\pi\alpha/3v_0$ in the denominator is a local field correction. The same factors are derived in the first term of $\mathscr{F}_{\mu\nu}$ in (4.685). The atomic polarizability α has dimensions of m^3, so the polarizability per unit volume α/v_0 is dimensionless. Often this combination itself is written as α. In any case, the dielectric function (4.686) may be derived from the correlation function $\mathscr{F}_{\mu\nu}$ in (4.685).

Now solve the Hamiltonian for the interaction between excitons and photons:

$$H = \sum_{\lambda \mathbf{k}} \omega_{\mathbf{k}} a^{\dagger}_{\lambda \mathbf{k}} a_{\lambda \mathbf{k}} + \frac{1}{2m} \sum_i \left[\mathbf{p}_i - \frac{e}{c} \mathbf{A}(\mathbf{r}_i) \right]^2 + \frac{e^2}{2} \sum_{ij} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{4.687}$$

This Hamiltonian can be rewritten in terms of the operators for excitons and photons. The electron–electron interaction is written in terms of the dipole–dipole approximation, as discussed above

$$\begin{aligned} H &= \sum_{\mathbf{k}\lambda} \omega_{\mathbf{k}} a^{\dagger}_{\mathbf{k}\lambda} a_{\mathbf{k}\lambda} + \sum_{\mathbf{k}\alpha} \omega_{\alpha} b^{\dagger}_{\alpha \mathbf{k}} b_{\alpha \mathbf{k}} + \frac{2\pi}{v_0} \sum_{\mu\nu\mathbf{k}} P_{\mu}(\mathbf{k}) T_{\mu\nu}(\mathbf{k}) P_{\nu}(\mathbf{k}) \\ &\quad - \frac{1}{\sqrt{v_0}} \sum_{\mu\mathbf{k}} J_{\mu}(\mathbf{k}) A_{\mu}(\mathbf{k}) + \frac{e^2 Z}{2m v_0} \sum_{\mu\mathbf{k}} A_{\mu}(\mathbf{k}) A_{\mu}(-\mathbf{k}) \\ J_{\mu}(\mathbf{k}) &= \frac{e}{m} \sum_{\alpha} p_{\mu}(\alpha)(b^{\dagger}_{\alpha\mathbf{k}} - b_{\alpha\mathbf{k}}) \\ p_{\mu}(\alpha) &= -i \int d^3 r \phi_m(\mathbf{r})^* \frac{\partial}{\partial x_{\mu}} \phi_n(\mathbf{r}) \end{aligned} \tag{4.688}$$

where the operator $J_{\mu}(\mathbf{k})$ is the current operator for the exciton system. The two operators J_{μ} and P_{μ} are simply related. The classical expression has the current as the time derivative of the polarization:

$$\mathbf{J} = \frac{\partial}{\partial t} \mathbf{P} \tag{4.689}$$

$$J_{\mu}(\mathbf{k}, \tau) = i \frac{\partial}{\partial \tau} P_{\mu}(\mathbf{k}, \tau) \tag{4.690}$$

This expression is also correct for the operators, which can be derived from the relation $\mathbf{p} = -im[\mathbf{r}, H]$ between momentum and position.

The photon Green's function is now obtained by the same procedure as used in Sec. 4.6.2. The two self-energy terms are evaluated as before. The first-order self-energy, which comes from the A^2 term, is evaluated with the same result [see (4.647)]:

$$\Pi^{(1)}_{\mu\nu} = \frac{e^2 Z}{m v_0} \delta_{\mu\nu} \tag{4.691}$$

The factor of Z is the number of electrons in each atom in unit volume v_0. It enters into the A^2 term because the Hamiltonian in (4.687) contains the summation over i, which is the summation over all electrons, i.e., the summation over each atom and then the summation over each electron on each atom. The plasma frequency in this system is then $4\pi\Pi^{(1)} = 4\pi e^2 Z/(m v_0) \equiv \omega_p^2$.

The other self-energy contribution comes from the $J \cdot A$ term in the Hamiltonian. In analogy with (4.652), this term in the self-energy arises from the correlation function:

$$\Pi^{(2)}_{\mu\nu} = \Phi_{\mu\nu}(\mathbf{k}, i\omega) = \frac{1}{v_0} \int_0^{\beta} d\tau e^{i\omega\tau} \langle T_{\tau} J_{\mu}(\mathbf{k}, \tau) J_{\nu}(-\mathbf{k}, 0) \rangle \tag{4.692}$$

This correlation function needs to be evaluated. It will be found by relating it to the correlation function for polarizability $\mathscr{F}_{\mu\nu}$ which is defined in (4.677) and solved in (4.685). This solution is the appropriate one for our case, since $\mathscr{F}_{\mu\nu}$ has been evaluated with the full exciton Hamiltonian, which includes all terms in the exciton–photon Hamiltonian (4.688)

except the $J \cdot A$ and A^2 terms. The exact solution is found for the photon self-energy, for the Hamiltonian (4.688), if we can relate $\Phi_{\mu\nu}$ to $\mathscr{F}_{\mu\nu}$. They are related by using the operator identity (4.690),

$$\Phi_{\mu\nu}(\mathbf{k}, i\omega) = -\frac{1}{\nu_0}\int_0^\beta d\tau e^{i\omega\tau}\left\langle T_\tau \frac{\partial}{\partial t}P_\mu(\mathbf{k}, \tau)\left[\frac{\partial}{\partial \tau'}P_\nu(-\mathbf{k}, \tau')\right]_{\tau'=0}\right\rangle$$

and then integrating by parts on the τ variable. The first integration by parts brings

$$\begin{aligned}\Phi_{\mu\nu}(\mathbf{k}, i\omega) = &-\frac{1}{\nu_0}\left\langle [P_\mu(\mathbf{k}, \beta) - P_\mu(\mathbf{k}, 0)]\left(\frac{\partial P_\nu}{\partial \tau'}\right)_{\tau'=0}\right\rangle \\ &-\frac{i\omega}{\nu_0}\int_0^\beta d\tau e^{i\omega\tau}\left\langle T_\tau P_\mu(\mathbf{k}, \tau)\left[\frac{\partial}{\partial \tau'}P_\nu(-\mathbf{k}, \tau')\right]_{\tau'=0}\right\rangle\end{aligned}$$

The argument of the correlation function is rewritten as

$$-\left\langle T_\tau P_\mu(\mathbf{k}, 0)\frac{\partial}{\partial t}P(-\mathbf{k}, -\tau)\right\rangle \tag{4.693}$$

and another integration by parts brings us to the expression

$$\begin{aligned}\Phi_{\mu\nu}(\mathbf{k}, i\omega) = &-\frac{1}{\nu_0}\left\langle [P_\mu(\mathbf{k}, \beta) - P_\mu(\mathbf{k}, 0)]\left(\frac{\partial P_\nu}{\partial \tau'}\right)_{\tau'=0}\right\rangle \\ &-\frac{i\omega}{\nu_0}\langle P_\mu(\mathbf{k}, 0)[P_\nu(-\mathbf{k}, -\beta) - P_\nu(-\mathbf{k}, 0)]\rangle \\ &+\frac{(i\omega)^2}{\nu_0}\int_0^\beta d\tau e^{i\omega\tau}\langle T_\tau P_\mu(\mathbf{k}, \tau)P_\nu(-\mathbf{k}, 0)\rangle\end{aligned}$$

The last term has the desired form of $(i\omega_n)^2\mathscr{F}_{\mu\nu}$. The first two terms, which are constants of integration, must be evaluated next. The second term is easy, since it is zero:

$$\begin{aligned}\langle P_\mu(\mathbf{k}, 0)[P_\nu(-\mathbf{k}, -\beta) - P_\nu(-\mathbf{k}, 0)]\rangle &= \mathrm{Tr}\{e^{-\beta H}[P_\mu(\mathbf{k})e^{-\beta H}P_\nu(-\mathbf{k})e^{\beta H} - P_\mu(\mathbf{k})P_\nu(-\mathbf{k})]\} \\ &= -\mathrm{Tr}\{e^{-\beta H}[P_\mu(\mathbf{k}), P_\nu(-\mathbf{k})]\} = 0\end{aligned} \tag{4.694}$$

where the cyclic properties of the trace are used to rearrange the first term. The first term in $\Phi_{\mu\nu}$ requires more care, since it is nonzero and makes an important contribution to the result:

$$\begin{aligned}R &= -\frac{1}{\nu_0}\left\langle [P_\mu(\mathbf{k}, \beta) - P_\mu(\mathbf{k}, 0)]\left(\frac{\partial P_\nu}{\partial \tau'}\right)_{\tau'=0}\right\rangle \\ &= -\frac{1}{\nu_0}\left\langle \left(\frac{\partial P_\nu}{\partial \tau'}\right)_{\tau'=0}P_\mu - P_\mu\left(\frac{\partial P_\nu}{\partial \tau'}\right)_{\tau'=0}\right\rangle = -\frac{i}{\nu_0}[P_\mu, J_\nu]\end{aligned}$$

The commutators of J and P can be evaluated from their definitions:

$$\begin{aligned}R &= -\frac{e^2}{im\nu_0}\sum_{\alpha\beta}X_\mu(\alpha)p_\nu(\beta)[b_{\alpha\mathbf{k}} + b^\dagger_{\alpha-\mathbf{k}}, b^\dagger_{\beta\mathbf{k}} - b_{\beta,-\mathbf{k}}] \\ &= -\frac{2e^2}{im\nu_0}\sum_\alpha X_\mu(\alpha)p_\nu(\alpha)\end{aligned} \tag{4.695}$$

To evaluate the remaining summation over α, rewrite the momentum matrix element as $ip_x/m = [x, H]$, so that this term is

$$\frac{1}{m}p_\nu(\alpha) = i\omega_\alpha X_\nu(\alpha) \tag{4.696}$$

$$R = -\frac{2e^2}{v_0}\sum_\alpha \omega_\alpha X_\mu(\alpha)X_\nu(\alpha) = -\frac{e^2}{mv_0}\delta_{\mu\nu}\sum_\alpha f_\alpha \tag{4.697}$$

$$f_\alpha = \frac{2\omega_\alpha X_\mu^2(\alpha)m}{\hbar} \tag{4.698}$$

The summation over oscillator strength f_α can be trivially evaluated by using the f-sum rule,

$$\sum_\alpha f_\alpha = Z \tag{4.699}$$

where Z is the number of electrons on each atom. These steps show that the first term in (4.694) is just

$$R = -\frac{e^2 Z}{mv_0}\delta_{\mu\nu} \tag{4.700}$$

$$\Pi_{\mu\nu}^{(2)} = \Phi_{\mu\nu} = -\frac{\omega_P^2}{4\pi}\delta_{\mu\nu} + (i\omega)^2\mathscr{F}_{\mu\nu} \tag{4.701}$$

The two self-energy expressions may be added,

$$4\pi(\Pi^{(1)} + \Pi^{(2)}) = 4\pi(i\omega)^2\mathscr{F}_{\mu\nu}(\mathbf{k}, i\omega) \tag{4.702}$$

The plasma frequency cancels from the result. The photon self-energy function is just proportional to the correlation function $\mathscr{F}_{\mu\nu}$ of the polarization operator. Earlier it was called the polarizability of the solid. It is given in (4.685). Next, solve the Dyson equation for the photon Green's function:

$$\mathscr{D}_{\mu\nu} = \mathscr{D}_{\mu\nu}^{(0)} + \sum_{\lambda\alpha}\mathscr{D}_{\mu\lambda}^{(0)}\Pi_{\lambda\alpha}\mathscr{D}_{\alpha\nu} \tag{4.703}$$

By using the fact that $\mathscr{D}_{\mu\nu}$ and $\mathscr{D}_{\mu\nu}^{(0)}$ both have the factor $(\delta_{\mu\nu} - k_\mu k_\nu/k^2)$, it is easy to show that the term $D(k_\mu k_\nu/k^2)$ in $\mathscr{F}_{\mu\nu}$ makes no contribution. The solution to (4.703) is

$$\mathscr{D}_{\mu\nu} = \frac{4\pi(\delta_{\mu\nu} - k_\mu k_\nu/k^2)}{(i\omega)^2\varepsilon(i\omega) - c^2k^2} \tag{4.704}$$

where the dielectric function is given in (4.686). The photon Green's function is shown to be governed by the transverse dielectric function $\varepsilon(i\omega)$, which has the Lorenz–Lorentz local field correction. The propagating normal modes are governed by the relationship $\omega^2\varepsilon(\omega) = \omega_k^2$, which gives the polariton modes. The quantity $\mathscr{D}_{\mu\nu}$ is the Green's function for the polariton, since it has the correct dispersion relation. This interpretation is proper, since the function has the true eigenfrequencies of the exciton–photon system.

Since both the exciton and photon system are bosons, the problem could have as easily been solved from the exciton viewpoint. Then the photons merely give rise to self-energy corrections for the exciton system. When calculating the self-energy of the exciton, with all photon effects included, the same dispersion relation is derived for polaritons. Since the photon and exciton systems have equal footing, the problem can be treated successfully from either starting point. The solution from the exciton system is assigned as a problem.

The only real difference between the phonon and exciton solutions, in Sec. 4.6.2 and this section, is the inclusion of the local field corrections for the excitons. They arise from the term in the Hamiltonian describing the dipole–dipole interaction between polarization modes on different atoms. Of course, the phonon system also has this term in the Hamiltonian, which should be included in the analysis.

Problems

1. For the hard sphere potential, show that the diagonal reaction matrix is

$$R_l(k,k) = \frac{\hbar^2}{2mk}\frac{j_l(ka)}{\eta_l(ka)} \tag{4.705}$$

Next show that the exact off-diagonal reaction matrix is

$$R_l(k',k) = \frac{\hbar^2}{2mk}\frac{j_l(k'a)}{\eta_l(ka)} \tag{4.706}$$

(Hint: Let the hard sphere be a finite step with height V_0, and solve for $V_0 \to \infty$.) What is the T matrix for the hard sphere?

2. Consider a one-dimensional harmonic chain of infinite length. Let one atom have a mass m different from the others of mass M. Solve exactly for the vibrational normal modes, and obtain the condition for the formation of a local mode of vibration. Does this happen when the impurity mass m is lighter or heavier than M?

3. Consider the vibrational modes of a three-dimensional solid. Let all atoms be alike except for one. Obtain the equation which determines the frequency of the normal mode. (a) Isotope effect: let the one atom have a different mass but the same force constants. (b) Impurity: let both mass and force constants be different.

4. For the two-dimensional electron gas where the particles have kinetic energy $\varepsilon_k = k^2/2m$, derive:

- The formula for the cross section in terms of the phase shifts. (Question: what are the dimensions of cross section in 2D?)
- The formula for the on-shell T-matrix in terms of phase shifts.
- The Friedel oscillations $\delta n(r)$ in terms of phase shifts.
- The change in the ground state energy due to the impurity in terms of phase shifts.

5. Let two particles move with the same constant velocity $\mathbf{v}$ in a phonon field. The Hamiltonian is

$$H = \mathbf{v}\cdot(\mathbf{p}_1 + \mathbf{p}_2) + \sum_{\mathbf{q}}\left[\omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} + \frac{M_{\mathbf{q}}}{\sqrt{\nu}} A_{\mathbf{q}}(e^{i\mathbf{q}\cdot\mathbf{R}_1} + e^{i\mathbf{q}\cdot\mathbf{R}_2})\right]$$

(a) Show that the interaction energy due to phonon exchange is ($\mathbf{r} = \mathbf{R}_1 - \mathbf{R}_2$)

$$V(\mathbf{r},\mathbf{v}) = -2\sum_{\mathbf{q}} \frac{M_{\mathbf{q}}^2}{\nu}\frac{\cos(\mathbf{q}\cdot\mathbf{r})}{\omega_{\mathbf{q}} - \mathbf{q}\cdot\mathbf{v}} \tag{4.707}$$

(b) For a Debye model with piezoelectric coupling, this expression may be approximated as

$$V(\mathbf{r}, \mathbf{v}) = -M_0 \int_{-\infty}^{\infty} \frac{dq}{2\pi} \int \frac{d\Omega_q}{4\pi} \frac{e^{i\mathbf{q}\cdot\mathbf{r}}}{c_s - \hat{\mathbf{q}}\cdot\mathbf{v}} \tag{4.708}$$

Evaluate this integral as a function of r and v.

6. Let the following Hamiltonian describe the phonon interaction with a collection of constant velocity particles: $H = H_0 + V$, where

$$H_0 = \sum_j \mathbf{v}_j \cdot \mathbf{p}_j + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} \tag{4.709}$$

$$V = \frac{1}{\sqrt{\nu}} \sum_{\mathbf{q}} M_{\mathbf{q}} A_{\mathbf{q}} \sum_j e^{i\mathbf{q}\cdot\mathbf{r}_j} \tag{4.710}$$

where $\mathbf{r}_j$ and $\mathbf{p}_j$ are the position and momentum of the particles. Evaluate

$$U(t) = \langle e^{iH_0 t} e^{-iHt} \rangle \tag{4.711}$$

where the average is over the phonon thermodynamics.

7. The coherent neutron scattering from the atoms in a solid is described by the correlation function

$$S(\mathbf{q}, t) = \frac{1}{N} \sum_{jl} \langle e^{i\mathbf{q}\cdot\mathbf{R}_j(t)} e^{-i\mathbf{q}\cdot\mathbf{R}_l(0)} \rangle \tag{4.712}$$

For a crystal, write $\mathbf{R}_j(t) = \mathbf{R}_j^{(0)} + \mathbf{Q}_j(t)$, where $\mathbf{Q}_j(t)$ is the displacement due to phonons. Find the exact result for this correlation function.

8. Derive a result for the second moment

$$f = \langle \omega^2 \rangle - \langle \omega \rangle^2 \tag{4.713}$$

for the spectral function (4.283). See (4.269) for definitions of $\langle \omega^n \rangle$.

9. Show that the free energy of the Fano–Anderson model can be written exactly as $F = F_0 + \delta F$ where

$$\beta F_0 = 2 \sum_k \ln[1 + e^{-\beta\varepsilon_k}] + \ln[1 + e^{-\beta\varepsilon_c}] \tag{4.714}$$

$$\delta F = \int_0^1 d\lambda \sum_{ip} \frac{\Sigma(ip)}{ip - \varepsilon_c - \lambda\Sigma(ip)} \tag{4.715}$$

Evaluate this expression at zero temperature, and show that it agrees with Fumi's theorem.

10. Solve for the density of state of the Bethe lattice with $z = 2$: Show that it is exactly equal to the results for a simple one-dimensional chain of sites.

11. Integrate the density of state $\rho(\Omega)$ for the Bethe lattice and show that the total number of states in the band is one per site, as expected.

12. Use perturbation theory to evaluate the Green's function for the Bethe lattice

$$\mathcal{G}(1, ip_n) = -\int_0^\beta d\tau e^{ip_n\tau} \langle T_\tau C_{j+1}(\tau) C_j^\dagger(0)\rangle \tag{4.716}$$

$$= \sum_{m=0} \frac{t^{2m+1}}{(ip_n + \mu)^{2m+2}} f_{2m+1}(z) \tag{4.717}$$

where $j+1$ is a near neighbor of j. Show that the Green's function has the expansion indicated in the second line. Find the first three functions f_1, f_3, f_5, where z is the number of neighbors. Can these be summed into a Dyson's type of equation?

13. Solve the Ising model for a one-dimensional chain. Start at one end of the chain, and start averaging over spins, and derive a self-consistent equation for the partition function.

14. Solve the Ising model for the Bethe lattice for the case of antiferro-magnetic interactions ($J < 0$). For $h = 0$ prove that it has the same transition temperature as the case of ferromagnetic ordering.

15. Solve the Tomonaga model when the interaction

$$U = \frac{1}{2L} \sum_k U_k [\rho(k)\rho(-k) + \sigma(k)\sigma(-k)] \tag{4.718}$$

is added to the original Hamiltonian (4.465). Diagonalize this new Hamiltonian, and use it to evaluate the Pauli spin susceptibility (4.523).

16. Consider the static density–density correlation function

$$c(\xi) = \frac{1}{L}\int_0^L dx \rho(x)\rho(x+\xi) = \frac{1}{L^2}\sum_k \cos(k\xi)\langle\rho(k)\rho(-k)\rangle$$

(a) Evaluate this for the noninteracting electron system at $T = 0$.
(b) Use the Tomonaga model to evaluate this for the interacting electron system.

17. In the spinless Luttinger model, compare the commutator $[\Psi_1(x), H_0]$ as obtained in the fermion representation and the boson representation.

18. Calculate the retarded density–density correlation function for the spinless Luttinger model:

$$\chi(p, i\omega) = \int_0^\beta d\tau e^{i\omega\tau} \langle T_\tau \rho(p, \tau)\rho(-p, 0)\rangle \tag{4.719}$$

$$\rho(p) = \rho_1(p) + \rho_2(p) \tag{4.720}$$

19. Interesting effects happen to the susceptibilities $\chi(q, \omega)$ when $q \sim 2k_F$. This wave vector region may be evaluated in the Luttinger model by taking operators which are the product of particle-1 and -2 fermions. For example, evaluate the correlation function $\chi(q, \omega)$ in the noninteracting spinless Luttinger model, where

$$\Lambda(q) = \int dx e^{-iqx} \Psi_{2s}^\dagger(x)\Psi_{1s}(x) \tag{4.721}$$

$$\chi(q, t) = \langle \Lambda(q, t)\Lambda^\dagger(q, 0)\rangle \tag{4.722}$$

20. Find the eigenvalue equation of the boson modes of a one-dimensional electron gas described by the spinless Luttinger model, which interacts with phonons. The ds are phonon operators:

$$H = v_F \sum_k k(a_{1s}^\dagger a_{1s} - a_{2s}^\dagger a_{2s}) + \sum_k \Omega_k d_k^\dagger d_k + \sum_{\text{all } q} \frac{M_q}{\sqrt{L}}(d_q + d_{-q}^\dagger)[\rho_1(q) + \rho_2(q)] \tag{4.723}$$

21. Write the Hamiltonian (4.641) in terms of harmonic oscillator coordinates for both the photons and the phonons. Solve it as a coupled oscillator problem.

22. Solve the polariton Hamiltonian (4.688) from the exciton viewpoint. Calculate the self-energies of the exciton, including those arising from the interaction with photons. You may simplify the problem by neglecting the local field term. Your answer should have the poles of the exciton Green's function at the polariton modes.

Chapter 5

Homogeneous Electron Gas

The use of diagram techniques in many-particle physics began in the early 1950s, soon after their introduction into field theory. Although these methods were applied to a variety of problems, some areas of work were more successful than others. The two areas which enjoyed early success were the homogeneous electron gas and the polaron problem. Later there were other successes such as the theories of superconductivity and superfluidity. However, the theory of the homogeneous electron gas, as it was initially understood, was worked out by many contributors during the period 1957–1958. They brought a variety of theoretical approaches to this problem, but all used diagrammatic techniques in some form. On the other hand, during the past twenty years, there has been achieved an understanding of electron–electron interactions in strongly correlated metals. This latter topic is covered in the next chapter.

5.1. EXCHANGE AND CORRELATION

The homogeneous electron gas is described by the Hamiltonian

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} C^{\dagger}_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \frac{1}{2\nu} \sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} \sum_{\mathbf{q}\neq 0} v_q C^{\dagger}_{\mathbf{k}+\mathbf{q}\sigma} C_{\mathbf{k}'-\mathbf{q},\sigma'} C_{\mathbf{k},\sigma'} C_{\mathbf{k}}\sigma$$

$$\varepsilon_{\mathbf{p}} = \frac{p^2}{2m} \tag{5.1}$$

$$v_q = \frac{4\pi e^2}{q^2} \tag{5.2}$$

which was derived in (1.164). The electrons are free particles, which mutually interact by Coulomb's law e^2/r. There are N_e electrons in a large volume ν, with an average density $n_0 = N_e/\nu$. A positive charge of density n_0 is spread uniformly through the volume ν. The positive background maintains the overall charge neutrality of the system. The homogeneous electron gas is also called the *jellium model* of a solid.

There are two electronic properties which will be evaluated in this section. First is the self-energy of an electron of momentum $\mathbf{p}$. Second, the properties of all the electrons will be averaged to obtain the total energy of the system. This average is done at zero temperature, which derives the ground state energy. The important quantity is the ground state energy per

particle E_g, which can depend only on the particle density $E_g(n_0)$. The total energy of the N_e-particle system is just $N_e E_g = E_T$, since surface effects are ignored.

The parameter r_s is universally used to describe the density of an electron gas,

$$\frac{4\pi n_0 a_0^3}{3} r_s^3 = 1 \tag{5.3}$$

where a_0 is the Bohr radius. In an electron gas with uniform density n_0, r_s is the radius in atomic units of the sphere which encloses one unit of electron charge. Thus r_s is small for a high-density electron gas and it is large for a low-density gas. Other properties of an electron gas may be expressed in terms of this parameter. The density may be related to the Fermi wave vector,

$$n_0 = 2\int \frac{d^3p}{(2\pi)^3} n_p = \frac{1}{\pi^2}\int_0^{k_F} p^2 dp = \frac{k_F^3}{3\pi^2} \tag{5.4}$$

so that the Fermi wave vector and energy are related to r_s,

$$\begin{aligned} k_F a_0 &= (3\pi^2 n_0)^{1/3} a_0 = \left(\frac{9\pi}{4}\right)^{1/3}\left(\frac{4\pi n_0 a_0^3}{3}\right)^{1/3} = \left(\frac{9\pi}{4}\right)^{1/3}\frac{1}{r_s} = \frac{1.9192}{r_s} \\ E_F &= \frac{\hbar^2 k_F^2}{2m} = (k_F a_0)^2\left(\frac{\hbar^2}{2ma_0^2}\right) = \frac{3.6832}{r_s^2} E_{ry} \end{aligned} \tag{5.5}$$

where $E_{ry} = 13.60\,\text{eV}$ will be the standard unit of energy. Similarly, the plasma frequency is

$$\hbar\omega_p = \hbar\left(\frac{4\pi e^2 n_0}{m}\right)^{1/2} = \left(\frac{12}{r_s^3}\right)^{1/2} E_{ry} = \frac{3.4641}{r_s^{3/2}} E_{ry} \tag{5.6}$$

In the homogeneous electron gas, the average kinetic energy of the electrons is going to be proportional to $E_F \approx \langle K.E.\rangle \sim k_F^2$, which, by dimensional analysis, is inversely proportional to the square of the characteristic length of the system, which is r_s. Therefore $\langle K.E.\rangle \propto 1/r_s^2$. Similarly, dimensional analysis suggests that the average Coulomb energy per particle will be e^2 divided by the characteristic length, or $\langle P.E.\rangle \propto 1/r_s$. When the electron gas has sufficiently high density, which is small r_s, the kinetic energy term will be larger than the potential energy term. In this case, the electrons will behave as free particles, since the potential energy is a perturbation on the dominant kinetic energy. In the high-density limit, the free-particle picture is expected to be valid. This limiting case will be investigated below. The kinetic energy and potential energy contributions will be calculated. The potential energy terms cannot be found exactly, but the result can be expressed as a power series—with logarithm terms—in the parameter r_s. This series should be accurate at small values of r_s. A term-by-term derivation of this series is derived below. Later, in Sec. 5.2, the other limit of low density is considered where the potential energy is larger than the kinetic energy.

5.1.1. Kinetic Energy

The first energy term is the kinetic energy. For a single particle it is $\varepsilon_k = k^2/2m$. The contribution to the ground state energy is obtained by summing over all the particles in the ground state:

$$E_{T,KE} = \sum_{\mathbf{p}\sigma} \varepsilon_p n_p = 2\nu \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{2m} n_p = \left(\frac{N_e}{n_0}\right) \frac{1}{2\pi^2 m} \int_0^{k_F} p^4 dp$$
$$= \frac{3}{5} \frac{\hbar^2 k_F^2}{2m} N_e = \tfrac{3}{5} E_F N_e \tag{5.7}$$

$$\tfrac{3}{5} E_F = \frac{2.2099}{r_s^2} E_{ry} \tag{5.8}$$

The average kinetic energy is $\frac{3}{5}E_F$, which is given in terms of r_s.

5.1.2. Hartree

All the remaining terms in the energy come from the Coulomb interaction between the particles. This contribution has not been evaluated exactly. Instead, approximate expressions are obtained by a variety of means. Our derivation follows that of Gell-Mann and Brueckner (1957). They start by examining the terms generated by ordinary perturbation theory. The first term which occurs is the Coulomb interaction between the electrons and the uniform positive background, which is called the *Hartree interaction*. In the model of the homogeneous electron gas, the time-averaged electron density is uniform throughout the system, as is the positive background. These equal and opposite charge densities exactly cancel, so that the net system is charge neutral. The Hartree energy is zero. That is, this energy is given by the equation

$$N_e E_0 = \frac{e^2}{2} \int \frac{d^3r_1 d^3r_2}{|\mathbf{r}_1 - \mathbf{r}_2|} [\rho_e(\mathbf{r}_1) - \rho_i(\mathbf{r}_1)][\rho_e(\mathbf{r}_2) - \rho_i(\mathbf{r}_2)] \tag{5.9}$$

but the ion and electron particle densities are $\rho_e = \rho_i = n_0$, and the contribution is zero. This fact has already been used in writing Eq. (5.1) by the omission of the $\mathbf{q} = 0$ term from the Coulomb interaction. The $\mathbf{q} = 0$ term is the direct Coulomb interaction among the electrons. This term is omitted because it is canceled by the direct interaction with the positive background.

5.1.3. Exchange

The Coulomb interactions in (5.1) provide other energy contributions in addition to the direct term. The Hartree term corresponds to the following pairing of the operators in (5.1) when the average value of H is evaluated:

$$\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} \langle C^\dagger_{\mathbf{k}+\mathbf{q},\sigma} C^\dagger_{\mathbf{k}'-\mathbf{q},\sigma'} C_{\mathbf{k}',\sigma'} C_{\mathbf{k}\sigma} \rangle \approx \delta_{\mathbf{q}=0} \sum_{\mathbf{k}\sigma} \langle C^+_{\mathbf{k},\sigma} C_{\mathbf{k},\sigma} \rangle \sum_{\mathbf{k}'\sigma'} \langle C^+_{\mathbf{k}',\sigma'} C_{\mathbf{k}',\sigma'} \rangle$$

Another way to pair the same operators is

$$\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} \langle C^\dagger_{\mathbf{k}+q,\sigma} C^\dagger_{\mathbf{k}'-\mathbf{q},\sigma'} C_{\mathbf{k}'\sigma'} C_{\mathbf{k}\sigma}\rangle \approx -\sum_{\mathbf{k}\mathbf{k}'\sigma\sigma'} \langle C^\dagger_{\mathbf{k}+\mathbf{q},\sigma} C_{\mathbf{k}'\sigma'}\rangle\langle C^\dagger_{\mathbf{k}'-\mathbf{q},\sigma'} C_{\mathbf{k}\sigma}\rangle$$
$$= -\sum_{\mathbf{k}\sigma} n_{\mathbf{k}+\mathbf{q}} n_{\mathbf{k}}$$

This arrangement requires that $\sigma' = \sigma$ and $\mathbf{k}' = \mathbf{k} + \mathbf{q}$. The minus sign is from exchanging the order of the fermion operators. This term is called the exchange energy, or the Fock energy. Retaining both terms is called *Hartree–Fock*. The exchange term was derived previously in (2.138). It gives a contribution to the energy of an individual electron, as well as a contribution to the ground state energy of the collection of electrons:

$$\Sigma_x(k) = -\frac{1}{\mathrm{v}} \sum_{\mathbf{q}} v_q n_{\mathbf{k}+\mathbf{q}} \tag{5.11}$$

$$E_{gx} = \frac{1}{2N_e} \sum_{\mathbf{k}\sigma} n_{\mathbf{k}} \Sigma_x(k) \tag{5.12}$$

The self-energy $\Sigma_x(k)$ depends only on the magnitude of the wave vector k of the particle and not upon its energy variable ip_n or ω. The wave vector integrals are elementary ($\mathbf{p} = \mathbf{k} + \mathbf{q}$, $\mathrm{v} = \cos\theta$):

$$\Sigma_x(k) = -\int \frac{d^3p}{(2\pi)^3} \frac{4\pi e^2}{|\mathbf{p} - \mathbf{k}|^2} n_p \tag{5.13}$$

$$= -\frac{e^2}{\pi} \int_0^{k_F} p^2 dp \int_{-1}^{1} \frac{d\mathrm{v}}{k^2 + p^2 - 2pk\mathrm{v}} \tag{5.14}$$

$$= -\frac{e^2}{\pi k} \int_0^{k_F} p dp \ln\left|\frac{k+p}{k-p}\right| \tag{5.15}$$

$$\Sigma_x(k) = -\frac{e^2 k_F}{\pi}\left(1 + \frac{1-y^2}{2y} \ln\left|\frac{1+y}{1-y}\right|\right) \tag{5.16}$$

$$y = \frac{k}{k_F} \tag{5.17}$$

A particle at the Fermi energy $k = k_F$ has $y = 1$ and

$$\Sigma_x(k_F) = -\frac{e^2 k_F}{\pi} \tag{5.18}$$

It is convenient to write

$$\Sigma_x(k) = \frac{e^2 k_F}{\pi} S(y) \tag{5.19}$$

$$S(y) = -\left(1 + \frac{1-y^2}{2y} \ln\left|\frac{1+y}{1-y}\right|\right) \tag{5.20}$$

where $S(y)$ is a function which gives the wave vector dependence of the exchange energy. This function is shown in Fig. 5.1. Its value at $y = 0$ is $S = -2$. In the vicinity of $y = 1$ it rises steeply and approaches zero at large values of y. At the Fermi energy the value is

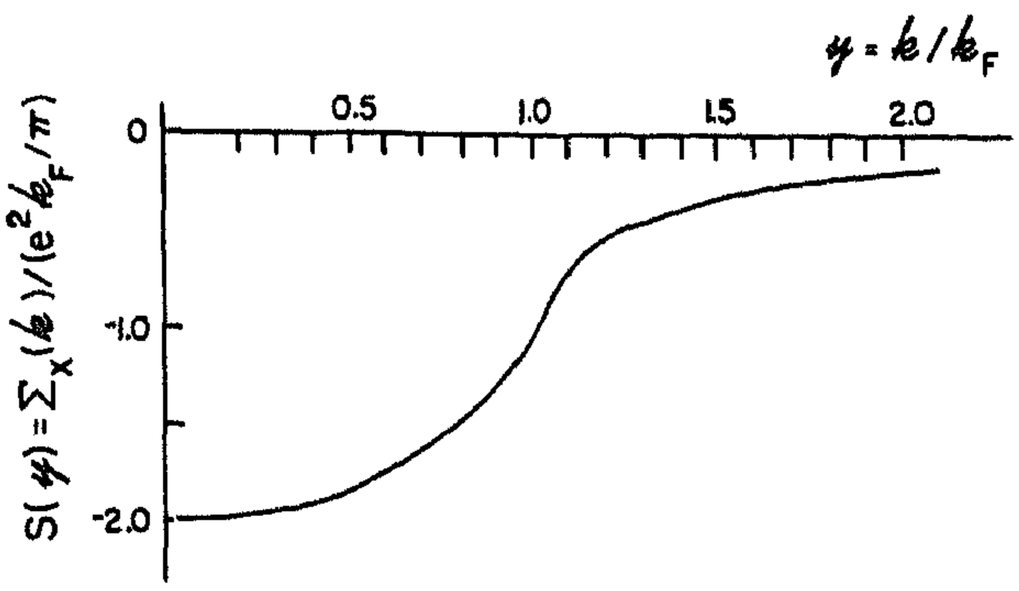


FIGURE 5.1 The unscreened exchange energy of the electron gas.

$S(1) = -1$. This point is interesting because the function $S(y)$ has an infinite slope. The derivative of $S(y)$ is

$$\frac{d}{dy}S(y) = \frac{1}{2y}\left(\frac{1+y^2}{y}\ln\left|\frac{1+y}{1-y}\right| - 2\right) \tag{5.21}$$

The derivative has a logarithmic divergence as $y \to 1$. This fact is interesting because it predicts that the effective mass is zero. The effective mass of a particle was defined in (3.160). Since the exchange self-energy is not frequency dependent, the effective mass is given by

$$\left(\frac{m}{m^*}\right) = 1 + \frac{\partial}{\partial \varepsilon_k}\Sigma_x(k) \tag{5.22}$$

$$= \frac{e^2 m}{\pi k}\frac{d}{dy}S(y) \tag{5.23}$$

$$= \frac{e^2 m}{2\pi k_F}\frac{1}{y^2}\left(\frac{1+y^2}{y}\ln\left|\frac{1+y}{1-y}\right| - 2\right) \tag{5.24}$$

which diverges at the Fermi energy $y \to 1$. If the inverse effective mass really diverged at the Fermi surface, it would have several observable consequences. The electron gas would be unstable at low temperatures, and the specific heat would diverge. Many metals have unusual properties at low temperatures, e.g., some become superconducting, while others become magnetic. However, simple metals such as the alkalis are stable at low temperatures, so that this exchange instability is regarded as being absent. In fact, an examination of further terms in the perturbation theory produces another divergence in the effective mass which exactly cancels the one due to exchange. The effective mass and specific heat are not divergent. This subject is pursued further in Sec. 5.8.

The exchange energy contribution to the ground state energy is obtained from (5.12). Summing over spins gives an expression which is easy to evaluate

$$E_{gx} = \frac{1}{N_e}\sum_{\mathbf{k}} n_{\mathbf{k}}\Sigma_x(k) = \frac{1}{n_0}\int \frac{d^3k}{(2\pi)^3} n_k \Sigma_x(k)$$

$$= -\frac{k_F^3}{2\pi^2 n_0}\left(\frac{e^2 k_F}{\pi}\right)\int_0^1 y^2 dy\left(1 + \frac{1-y^2}{2y}\ln\left|\frac{1+y}{1-y}\right|\right)$$

$$E_{gx} = -\frac{3}{4}\frac{e^2 k_F}{\pi} \tag{5.25}$$

The average exchange energy per electron is $\frac{3}{2}$ of the value at the Fermi energy. The additional factor of $\frac{1}{2}$ enters the ground state energy to account for the fact that the exchange energy is a

pair interaction. In terms of the parameter r_s the total ground state exchange energy per electron is

$$E_{gx} = -\frac{3}{2\pi}(k_F a_0)\left(\frac{e^2}{2a_0}\right) = -\frac{0.9163}{r_s} \tag{5.26}$$

where the earlier result for $(k_F a_0)$ was used in the final expression.

So far two terms have been found for the energy of the particle,

$$E(k) = \frac{\hbar^2 k^2}{2m} + \Sigma_x(k) + \cdots \tag{5.27}$$

The corresponding two terms for the ground state energy per particle are,

$$E_g = \frac{2.2099}{r_s^2} - \frac{0.9163}{r_s} + \cdots \tag{5.28}$$

The ground state energy has the appearance of a power series, in increasing powers of r_s. Although it is usually unsafe to extrapolate from just two terms, in fact E_g is a series in r_s. The next term will be of order $O(r_s^0)$. The zeroth power could be interpreted as either a constant or as $\ln(r_s)$. In fact, both of these terms are present. The series has the form

$$E_g = \frac{2.2099}{r_s^2} - \frac{0.9163}{r_s} - 0.094 + 0.0622\ \ln(r_s) + \cdots \tag{5.29}$$

where the last two terms were first given by Gell-Mann and Brueckner (1957). This result has since been obtained by a variety of perturbation techniques. The Gell-Mann and Brueckner derivation was by Rayleigh–Schrödinger perturbation theory. Alternate derivations have been presented by Sawada *et al.* (1957), Hubbard (1957), Nozieres and Pines (1958), and Quinn and Ferrell (1958). Our derivation follows Quinn and Ferrell.

The above energy terms comprise the Hartree–Fock theory. It is defined to be the kinetic energy, the Hartree energy which is zero, and the exchange energy. Wigner and Seitz (1933, 1934) suggested the name *correlation energy* to mean the energy terms beyond Hartree–Fock. The name is applied both to the additional energy terms in the self-energy of an electron of wave vector **k**, and to the ground state energy obtained by averaging over all of the electrons. The total ground state energy per particle is written as

$$E_g = \frac{2.2099}{r_s^2} - \frac{0.9163}{r_s} + E_c \tag{5.30}$$

where the correlation energy E_c needs to be determined. Of course, some terms are known for its power series in r_s:

$$E_c = -0.094 + 0.0622\ln(r_s) + O(r_s) \tag{5.31}$$

This result is accurate in the limit of $r_s \to 0$. There is some uncertainty regarding the maximum value of r_s for which these few terms provide an accurate description. The radius of convergence of the power series is about $r_s \leq 1$. Actual metals have values of r_s up to about 6, and this series does not give sensible numbers at these low densities.

The term correlation energy is often applied to other quantities besides the total ground state energy. For example, the correlation energy of a particle of wave vector k are those terms beyond Hartree–Fock:

$$E(k) = \frac{\hbar^2 k^2}{2m} + \Sigma_x(k) + \Sigma_c(k, ik_n) \tag{5.32}$$

The self-energy from correlation $\Sigma_c(k, ik_n)$ depends upon the particle energy ik_n. The energy and wave vector can be averaged to obtain the contribution E_c to the ground state correlation energy.

5.1.4. Seitz's Theorem

The derivation of Quinn and Ferrell (1958) does calculate the correlation energy per particle. They then average over electron states to obtain the correlation contribution to the ground state energy. This averaging is based on a theorem of Seitz (1940) which relates the ground state energy to the chemical potential. The chemical potential is defined as the energy it takes to add or remove an electron from the material. It is the energy which divides the empty from the occupied states at zero temperature. Of course, it is just the Fermi energy of the metal. The chemical potential is the energy of an electron of momentum k_F:

$$\mu = \frac{\hbar^2 k_F^2}{2m} + \Sigma_x(k_F) + \text{Re}\ \Sigma_c(k_F, 0) \tag{5.33}$$

where $ik_n = 0$ is the chemical potential. The chemical potential μ is only a function of the density of the electron n_0. The theorem of Seitz is that

$$\mu(n_0) = \frac{d}{dn_0}[n_0 E_g(n_0)] = E_g + n_0 \frac{dE_g}{dn_0} \tag{5.34}$$

The proof is based on the definition of the chemical potential. It is the energy difference between a system with N_e particles and one with $N_e + 1$:

$$\mu = E_T(N_e + 1) - E_T(N_e) \tag{5.35}$$

For an N_e-particle system, the total energy is $E_T = N_e E_g$. For fixed volume ν, an $(N_e + 1)$-particle system has a density $(1/\nu)(N_e + 1) = n_0 + 1/\nu$. The total energy of the $(N_e + 1)$-particle system is

$$\begin{aligned} E_T(N_e + 1) &= (N_e + 1)E_g(n_0 + 1/\nu) = (N_e + 1)\left(E_g(n_0) + \frac{1}{\nu}\frac{dE_g}{dn_0}\right) \\ &= N_e E_g + E_g(n_0) + n_0 \frac{dE_g}{dn_0} + O\left(\frac{1}{\nu}\right) \end{aligned} \tag{5.36}$$

The difference between this expression and $N_e E_g$ is the chemical potential and gives the assertion in (5.34). In proving this theorem, the volume ν is kept fixed, as is the amount of positive charge. The $(N_e + 1)$-particle system has a slight charge imbalance, since it has one more unit of electron charge than positive charge. However, the extra charge is a negligible contribution to the energy. For example, if a body of average dimension L is uniformly charged with one unit of charge, then its Coulomb energy is of order e^2/L. But L for a macroscopic body is $L \approx 1$ cm, and e^2/L is 10^{-7} eV, which is negligible. For very small particles, where the dimension L is nanometers, the charging energy e^2/L is important.

The chemical potential is the negative of the work function. It is the energy required to remove an electron from the solid and take it to infinity with zero kinetic energy. If this vacuum level is called zero, then the measured work function just gives the negative of the value of the chemical potential. However, there is a surface correction to the work function, or

chemical potential, but not to the volume part of the ground state energy per particle. That is, write the total energy as

$$E_T = N_e E_g + A E_S \tag{5.37}$$

where A is the total surface area and E_S is the energy per unit surface area. For macroscopic bodies, E_g does not depend on the surface area. However, μ does have a term which depends on the surface—actually on the surface dipole layer. Write it as $\mu = \mu_B + \Delta\mu$, where $\Delta\mu$ is the contribution from the surface dipole layer and μ_B is from the bulk terms. The surface contribution $\Delta\mu$ to the work function was first recognized by Bardeen (1936) and calculated accurately by Lang and Kohn (1971). The theorem of Seitz actually just refers to μ_B.

The relationship (5.34) is valid term by term in the energy series. The Hartree–Fock terms may be expressed in powers of n_0:

$$\begin{aligned} E_{g,HF} &= \frac{2.2099}{r_s^2} - \frac{0.9163}{r_s} = \tfrac{3}{5} X n_0^{2/3} - \tfrac{3}{4} Y n_0^{1/3} \\ \tfrac{3}{5} X n_0^{2/3} &= \frac{2.2099}{r_s^2} = \frac{3}{5}\frac{E_F}{E_{ry}} = \tfrac{3}{5}(3\pi^2 n_0 a_0^3)^{2/3} \\ \tfrac{3}{4} Y n_0^{1/3} &= \frac{0.9163}{r_s} = \frac{3}{2\pi}(3\pi^2 n_0 a_0^3)^{1/3} \end{aligned} \tag{5.38}$$

Applying the theorem for the chemical potential gives

$$\mu_{B,HF} = \frac{d}{dn_0}(n_0 E_{g,HF}) = X n_0^{2/3} - Y n_0^{1/3} = E_F - \frac{e^2 k_F}{\pi} \tag{5.39}$$

The theorem correctly gives that the average kinetic energy is $\frac{3}{5}$ of the value at the Fermi energy, and the average exchange energy is $\frac{3}{4}$ the value at the Fermi energy. These results follow from their respective powers of n_0: $\frac{2}{3}$ and $\frac{1}{3}$.

The correlation energy for the ground state is a series in r_s or $n_0^{1/3}$. The theorem may be applied to this series on a term-by-term fashion:

$$\begin{aligned} E_c &= A + B\ln(r_s) + C r_s + \cdots \\ &= A - \tfrac{1}{3} B \ln\left(\frac{4\pi n_0 a_0^3}{3}\right) + C\left(\frac{3}{4\pi n_0 a_0^3}\right)^{1/3} + \cdots \end{aligned} \tag{5.40}$$

$$\mu_c = \frac{d}{dn_0}(n_0 E_c) = A - \tfrac{1}{3} B + B\ln(r_s) + \tfrac{2}{3} C r_s + \cdots \tag{5.41}$$

Quinn and Ferrell used this theorem to make a backward deduction. They first found μ_c by calculating the self-energy terms of an electron at the Fermi energy. Then the above relationships may be used to find the correlation energy for the ground state energy. The logarithm term has the same coefficient B for the two energy terms. Once this term is determined, the constant term for the ground state energy is obtained by adding $B/3$ to the constant term found for the chemical potential.

The self-energy of an electron, from Coulomb interactions, may be calculated by the methods described in Chapters 2 and 3. The self-energy function has an infinite number of terms. The basic procedure is to start examining some low-order terms and to deduce which terms contribute to the constant and $\ln(r_s)$ terms. The term *low-order* refers to the order of the self-energy diagram, which is the number of internal Coulomb lines.

5.1.5. $\Sigma^{(2a)}$

The exchange energy graph is shown as the first diagram in Fig. 5.2. It is the only contribution with one Coulomb line. The correlation energy is the sum of all contributions with two or more Coulomb lines. There are three diagrams, or self-energy terms, with two Coulomb lines. One of these is shown in Fig. 5.2(a). By using the rules for constructing diagrams in Sec. 3.4, this self-energy contribution is

$$\Sigma^{(2a)}(k) = \frac{1}{\nu^2\beta^2}\sum_{\mathbf{qq}'} v_q v_{q'} \sum_{iq_n iq_{n'}} \mathscr{G}^{(0)}(k+q)\mathscr{G}^{(0)}(k+q')\mathscr{G}^{(0)}(k+q+q')$$

The first summation over frequencies is given in (3.217):

$$\frac{1}{\beta}\sum_{iq_n} \mathscr{G}^{(0)}(k+q)\mathscr{G}^{(0)}(k+q+q') = \frac{n_F(\xi_{\mathbf{k+q}}) - n_F(\xi_{\mathbf{k+q+q'}})}{iq_{n'} + \xi_{\mathbf{k+q}} - \xi_{\mathbf{k+q+q'}}} \tag{5.42}$$

The second summation requires a new derivation, since one energy denominator has odd multiples of $i\pi/\beta$, and the other has even multiples. The answer is

$$\frac{1}{\beta}\sum_{iq_{n'}} \frac{\mathscr{G}^{(0)}(k+q')}{iq_{n'} + \xi_{\mathbf{k+q}} - \xi_{\mathbf{k+q+q'}}} = -\frac{n_B(\xi_{\mathbf{k+q+q'}} - \xi_{\mathbf{k+q}}) + n_F(\xi_{\mathbf{k+q'}})}{ik_n + \xi_{\mathbf{k+q+q'}} - \xi_{\mathbf{k+q}} - \xi_{\mathbf{k+q'}}}$$

where one occupation function is a boson distribution and the other fermion. The boson distribution may be eliminated by using the identity

$$[n_F(\xi_{\mathbf{k+q}}) - n_F(\xi_{\mathbf{k+q+q'}})]n_B(\xi_{\mathbf{k+q+q'}} - \xi_{\mathbf{k+q}}) = n_F(\xi_{\mathbf{k+q+q'}})[1 - n_F(\xi_{\mathbf{k+q}})]$$

which gives the final self-energy term:

$$\begin{aligned}\Sigma^{(2a)}(k) = -\frac{1}{\nu^2}\sum_{\mathbf{qq}'} &\frac{v_q v_{q'}}{ik_n + \xi_{\mathbf{k+q+q'}} - \xi_{\mathbf{k+q'}} - \xi_{\mathbf{k+q}}} \\ &\times \{n_F(\xi_{\mathbf{k+q'}})[n_F(\xi_{\mathbf{k+q}}) - n_F\xi_{\mathbf{k+q+q'}})] \\ &+ n_F(\xi_{\mathbf{k+k+q'}})[1 - n_F(\xi_{\mathbf{k+q}})]\}\end{aligned} \tag{5.43}$$

The self-energy of a particle is needed on the Fermi surface. Set $k = k_F$ and $ik_n = \xi_{k_F} = k_F^2/2m - \mu = 0$. The terms in the energy denominator largely cancel,

$$\xi_{\mathbf{k}} + \xi_{\mathbf{k+q+q'}} - \xi_{\mathbf{k+q}} - \xi_{\mathbf{k+q'}} = \frac{\mathbf{q}\cdot\mathbf{q}'}{m} \tag{5.44}$$

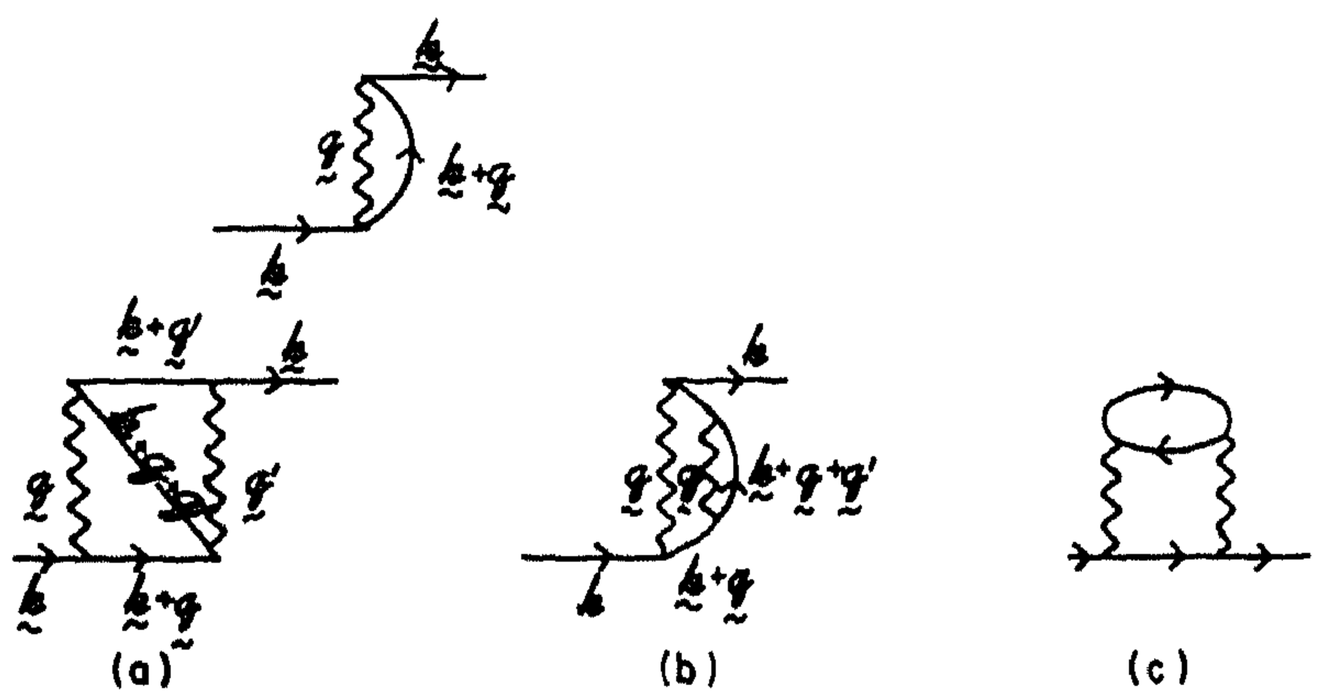


FIGURE 5.2

The self-energy is

$$\Sigma^{(2a)}(k_F, 0) = -\frac{(4\pi e^2)^2 m}{(2\pi)^6} \int \frac{d^3q}{q^2} \int \frac{d^3q'}{(q')^2} \frac{1}{\mathbf{q}\cdot\mathbf{q}'} \times \{n_F(\xi_{\mathbf{k}+\mathbf{q}'})[n_F(\xi_{\mathbf{k}+\mathbf{q}}) - n_F(\xi_{\mathbf{k}+\mathbf{q}+\mathbf{q}'}] + \cdots\} \tag{5.45}$$

This integral is evaluated by using dimensionless units. All wave vectors are normalized to the Fermi wave vector $\mathbf{x} = \mathbf{q}/k_F$, $\mathbf{y} = \mathbf{q}'/k_F$, $\hat{z} = \mathbf{k}_F/k_F$:

$$\Sigma^{(2a)} = -\frac{E_{ry}}{2\pi^4} \int \frac{d^3x}{x^2} \int \frac{d^3y}{y^2} \frac{1}{\mathbf{x}\cdot\mathbf{y}} \{n_F(\hat{z}+\mathbf{y})[n_F(\hat{z}+\mathbf{x}) - n_F(\hat{z}+\mathbf{x}+\mathbf{y})] + n_F(\hat{z}+\mathbf{x}+\mathbf{y})[1 - n_F(\hat{z}+\mathbf{x})]\} \tag{5.46}$$

The quantity on the right is independent of electron density. The integral is convergent and gives a nonzero result. It contributes to the constant term A in (5.41). Since Seitz's theorem is true for each term, this integral must equal the same constant as found for the equivalent term for E_c. The similar contribution to E_c was evaluated by Onsager *et al.* (1966) and found to be $\Sigma^{(2a)} = \frac{1}{3}\ln(2) - (3/2\pi^2)\zeta(3) = 0.0436$.

5.1.6. $\Sigma^{(2b)}$

The second self-energy term involving two Coulomb lines is shown in Fig. 5.2(b). This contribution is omitted because it is part of a sum of diagrams which gives zero. Consider the summation shown in Fig. 5.3(a). Each term has one more exchange diagram. All terms may be summed by evaluating the exchange energy with an electron Green's function in the self-energy which includes the exchange energy. This summation is given by the self-energy

$$\Sigma_x'(k) = -\frac{1}{\nu\beta} \sum_{\mathbf{q}, ik_n} v_q \mathscr{G}(\mathbf{k}+\mathbf{q}, ik_n) \tag{5.47}$$

$$\mathscr{G}(\mathbf{k}+\mathbf{q}, ik_n) = \frac{1}{ik_n - \xi_{\mathbf{k}+\mathbf{q}} - \Sigma_x(\mathbf{k}+\mathbf{q})} \tag{5.48}$$

where the Green's function has a self-energy due to exchange. Since the self-energy $\Sigma_x(k)$ does not depend on frequency, the frequency summation of the Green's function yields the simple number operator as in (3.219),

$$\frac{1}{\beta} \sum_{ik_n} \mathscr{G}(\mathbf{k}+\mathbf{q}, ik_n) = n_F[\xi_{\mathbf{k}+\mathbf{q}} + \Sigma_x(\mathbf{k}+\mathbf{q})] = \frac{1}{e^{\beta(\xi+\Sigma)} + 1}$$

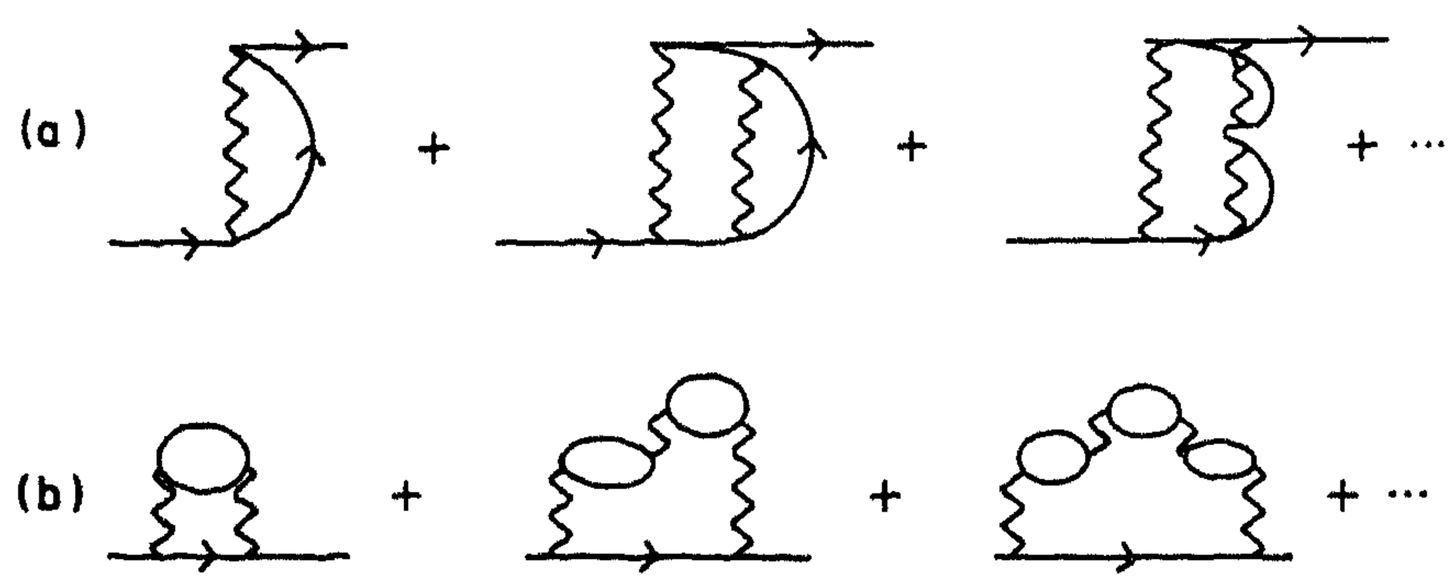


FIGURE 5.3

and the self-energy is

$$\Sigma_x'(k) = -\frac{1}{\nu}\sum_{\mathbf{q}} v_q n_F[\xi_{\mathbf{k}+\mathbf{q}} + \Sigma_x(\mathbf{k}+\mathbf{q})] = -\frac{1}{\nu}\sum_{\mathbf{p}} v_{\mathbf{p}-\mathbf{k}} n_F[\xi_{\mathbf{p}} + \Sigma_x(\mathbf{p})] \tag{5.49}$$

At zero temperature the electron distribution function $n_F[\xi_{\mathbf{p}} + \Sigma_x(\mathbf{p})]$ is just a step function which equals unity for electrons beneath the chemical potential and zero for those above. This step function must also be normalized so that the electron density is still n_0, as in

$$n_0 = 2\int \frac{d^3p}{(2\pi)^3} n_F[\xi_p + \Sigma_x(p)] \tag{5.50}$$

The effect of the exchange energy $\Sigma_x(p)$ in the argument of n_F is to change the chemical potential. However, the Fermi wave vector k_F is the same, even after the exchange energy has been included in the occupation function. The Fermi wave vector determines the density n_0, which is unchanged by the interactions. The addition of the exchange energy must be canceled by an equal change in chemical potential. The result is that at zero temperature,

$$\lim_{T\to 0} n_F[\xi_k + \Sigma_x(k)] = \Theta(k_F - k) \tag{5.51}$$

At zero temperature the Fermi distribution is still a step function at $k = k_F$. The exchange energy $\Sigma_x'(k)$ in (5.49) is exactly equal to the one calculated earlier. In Fig. 5.3(a), the summation of all terms just yields the value of the first term alone. All of the subsequent terms in that series sum to zero, because of a shift in the chemical potential.

The evaluation of these self-energy terms has become more subtle. It is not just a straightforward evaluation of terms in a series. One must understand each term in order to gauge its actual contribution. Further examples of this behavior are found in the other second-order term.

The exchange self-energy has a remarkable effect upon the zero temperature electron distribution. It leaves k_F unchanged. This simple result is a consequence of the feature that $\Sigma_x(k)$ is independent of frequency. Most self-energy terms depend on frequency and so have a larger effect upon the wave vector distributions: recall Problem 4 in Chapter 3.

5.1.7. $\Sigma^{(2c)}$

The third self-energy with two Coulomb lines is shown in Fig. 5.2(c). This self-energy diagram may be written as

$$\Sigma^{(2c)}(k) = -\frac{1}{\nu\beta}\sum_{\mathbf{q},iq_m} v_q^2 P^{(1)}(\mathbf{q}, iq_m)\mathscr{G}^{(0)}(\mathbf{k}+\mathbf{q}, ik_n + iq_m) \tag{5.52}$$

The closed fermion loop gives the polarization diagram $P^{(1)}(\mathbf{q}, iq_m)$. Although one could proceed with the evaluation of this self-energy, it obviously has one drawback. The wave vector integral appears to diverge at small values of q, when $\nu \to \infty$, because of the factor v_q^2

$$\int \frac{d^3q}{q^4} \Rightarrow \int_0 \frac{dq}{q^2} \to \infty \tag{5.53}$$

Indeed it does diverge, so this self-energy term is infinite. The divergence is removed by summing a series of self-energy diagrams. This series is shown in Fig. 5.3(b). Each term has

one more polarization bubble and one more Coulomb line—hence an additional factor of $v_q P^{(1)}$. The summation of these terms gives the simple series

$$\begin{aligned}\Sigma^{(3b)}(k) &= -\frac{1}{\nu\beta}\sum_q v_q \mathscr{G}^{(0)}(k+q) \\ &\quad \times \{v_q P^{(1)}(q) + [v_q P^{(1)}(q)]^2 + [v_q P^{(1)}(q)]^3 + \cdots\} \\ &= -\frac{1}{\nu\beta}\sum_q v_q \mathscr{G}^{(0)}(k+q)\frac{v_q P^{(1)}(q)}{1 - v_q P^{(1)}(q)}\end{aligned} \tag{5.54}$$

The summation of terms shown in Fig. 5.3 is called the *random phase approximation* (RPA). The "approximation" in RPA is that this series of terms is used to represent the entire answer. For example, the RPA result for the electron correlation energy uses Fig. 5.3(b) or (5.54). RPA ignores other terms, such as (5.46), which was found from the diagram in Fig. 5.2(a). RPA ignores many more terms in each higher order of perturbation series.

The denominator in (5.54) is a very important quantity because it is very often used in calculations. It is the RPA approximation to the dielectric function and is defined as

$$\varepsilon_{\mathrm{RPA}}(q) = 1 - v_q P^{(1)}(q) \tag{5.55}$$

The self-energy series may be written as

$$\Sigma^{(3b)}(k) = -\frac{1}{\nu\beta}\sum_q v_q \mathscr{G}^{(0)}(k+q)\left[\frac{1}{\varepsilon_{\mathrm{RPA}}(q)} - 1\right] \tag{5.56}$$

The "-1" term in the brackets is easy to evaluate. It is

$$\frac{1}{\nu\beta}\sum_q v_q \mathscr{G}^{(0)}(k+q) = \frac{1}{\nu}\sum_{\mathbf{q}} v_q n_{\mathbf{k}+\mathbf{q}} = -\Sigma_x(k) \tag{5.57}$$

which is just the negative of the exchange energy. Adding the exchange energy to the above result gives a term containing the RPA dielectric function:

$$\Sigma_{\mathrm{RPA}}(q) = \Sigma^{(3b)}(k) + \Sigma_x(k) = -\frac{1}{\nu\beta}\sum_q v_q \frac{\mathscr{G}^{(0)}(k+q)}{\varepsilon_{\mathrm{RPA}}(q)} \tag{5.58}$$

The summation of these two terms is defined as the RPA self-energy. It is frequently used to evaluate the self-energy of the electron (see Lundqvist, 1969). It will be discussed in Sec. 5.8. The exchange energy is logically included as the first term in the series shown in Fig. 5.3(b). The RPA self-energy, defined above, is also called the *screened exchange energy*. This self-energy has one feature which is worth mentioning now, although it will be shown later in Sec. 5.8: the inverse effective mass no longer diverges at the Fermi surface. A detailed evaluation of this self-energy shows that the coefficient $B = (2/\pi^2)(1 - \ln 2) = 0.0622$.

5.1.8. High-Density Limit

The first two terms in the series for the ground state energy in the RPA are

$$E_{c,\mathrm{RPA}} = -0.142 + 0.0622 \ln r_s + O(r_s, r_s \ln(r_s)) \tag{5.59}$$

Adding the Onsager result, for the second-order diagram $\Sigma^{(2a)}$ [see (5.46)], gives the result of Gell-Mann and Brueckner:

$$E_c = -0.094 + 0.0622 \ln r_s + \cdots + \tag{5.60}$$

Another term in the series has been obtained by Carr and Maradudin (1964):

$$E_c = -0.094 + 0.0622 \ln r_s + 0.018 r_s \ln r_s + a r_s + O(r_s^2) \tag{5.61}$$

They were unable to obtain all the terms contributing to the coefficient a of the linear term in r_s. They obtained all its contributions except one, and they were able to put limits on its value.

Carr and Maradudin also discussed the convergence of this series at large values of r_s. Because of the term $\ln r_s$, this series is not analytic as $r_s \to 0$, which is the limit of very high densities. But the series is accurate in this limit. Metallic densities are roughly $1.8 < r_s < 6$. It is an important question whether the formula (5.61) is valid for these values of r_s. Figure 5.4 shows a plot of the correlation energy found by Carr and Maradudin. Only the terms shown in (5.61) have been plotted ($a = 0$). It diverges at small r_s, which is just the effect of the $\ln(r_s)$ term. It is negative up to values about $r_s \approx 2.5$, where it crosses the axis and becomes positive. Now there are simple arguments which indicate that the correlation energy must be negative. These arguments are given below in great detail. They show that the perturbation formula for E_c is not valid in the range of metallic densities.

Large values of r_s correspond to the low-density electron gas. Wigner (1934) showed that the low-density electron gas behaved very differently from the high-density electron gas. At low density, the electrons become localized, and the calculation of the correlation energy becomes quite different. The low-density limit of Wigner is discussed in the next section. The correlation energy is determined for all values of r_s by extrapolation procedures between the low- and high-density limits. These extrapolation procedures provide a formula for the correlation energy which is valid in the region of metallic densities.

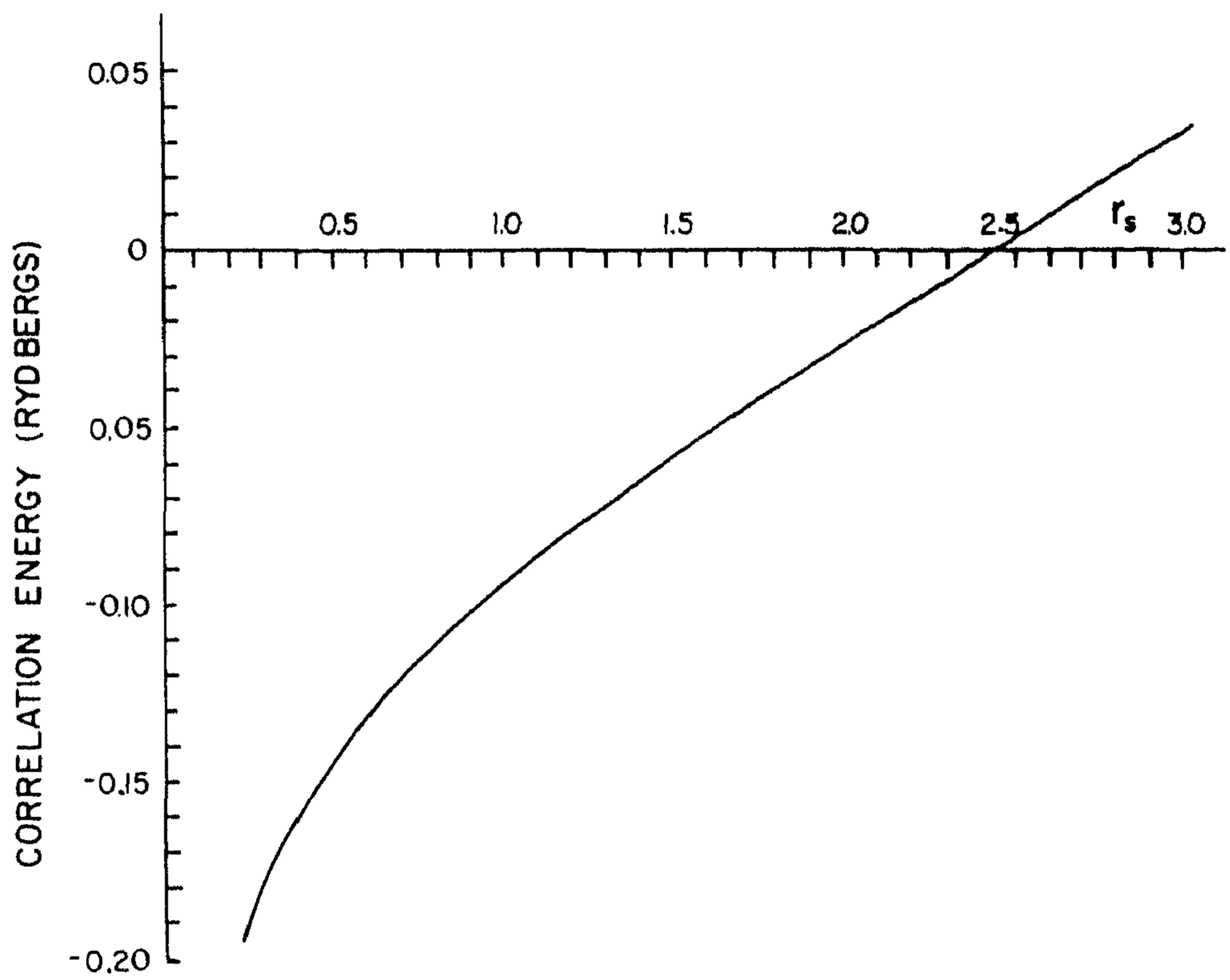


FIGURE 5.4 The correlation energy of the electron gas as given by the expansion of Carr and Maradudin.

5.1.9. Pair Distribution Function

The correlation energy is the improvement in the ground state energy beyond the Hartree–Fock approximation. The correlation energy may be understood, at least conceptually, only after understanding Hartree–Fock. It is useful to examine an important property of the Hartree–Fock theory: the *pair distribution function* $g(r)$, which was introduced in Sec. 1.6. The pair distribution function $g(r)$ is the probability per unit volume that an electron is at $\mathbf{r}$ if there is already one at $r = 0$. It depends on spin, so there are two different pair distribution functions in an unmagnetized electron gas: $g_{\uparrow\uparrow}(r) = g_{\downarrow\downarrow}(r)$ and $g_{\uparrow\downarrow}(r) = g_{\downarrow\uparrow}(r)$. In the notation $g_{ss'}(r)$ the first spin index s is for the central electron, while the second s' is for the electrons at $\mathbf{r}$. Thus $g_{\downarrow\uparrow}(r)$ is the probability that a spin-up electron is at $\mathbf{r}$ if a spin-down electron is at $r = 0$. Of course, the electrons are not fixed, and they are usually moving rapidly. These pair distribution functions are averages for the moving particles. Even the electron at $r = 0$ is not fixed, so that this reference point moves with the electron.

The pair distribution functions are rather easy to calculate in the Hartree–Fock approximation. Here the N-particle wave function is a Slater determinant,

$$\Psi_{\lambda_1\lambda_2\cdots\lambda_N}(r_1\cdots r_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_{\lambda_1}(r_1) & \phi_{\lambda_1}(r_2) & \cdots & \phi_{\lambda_1}(r_N) \\ \phi_{\lambda_2}(r_1) & \phi_{\lambda_2}(r_2) & \cdots & \phi_{\lambda_2}(r_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{\lambda_N}(r_1) & \phi_{\lambda_N}(r_2) & \cdots & \phi_{\lambda_N}(r_N) \end{vmatrix} \tag{5.62}$$

where the λ_j are the quantum numbers which describe the states, and there is one wave function for every occupied electron state. The square of this many-particle wave function is the N-particle density matrix. The pair distribution function is given by the two-particle density matrix. It is obtained by integrating the N-particle density over all but two space coordinates. A similar expectation value is taken over the spin variables:

$$g_{ss'}(\mathbf{r}_1, \mathbf{r}_2) = \mathrm{v}^2 \int d^3r_3 d^3r_4 \cdots d^3r_N |\Psi_{\lambda_1\lambda_2\cdots\lambda_N}(r_1\cdots r_N)|^2 \tag{5.63}$$

If the one-electron orbitals $\phi_{\lambda_j}(r)$ are assumed orthogonal, then this integration just yields the sum over all possible pair wave functions:

$$g_{ss'}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\mathrm{v}^2}{N(N-1)} \sum_{\lambda_i\lambda_j} \begin{vmatrix} \phi_{\lambda_i}(r_1) & \phi_{\lambda_i}(r_2) \\ \phi_{\lambda_j}(r_1) & \phi_{\lambda_j}(r_2) \end{vmatrix}^2 \tag{5.64}$$

The sum over λ_i, λ_j is over all occupied states, so each pair is summed twice. For the homogeneous electron gas, the orbitals must describe plane waves,

$$\phi_\lambda(\mathbf{r}) = \frac{\chi_s}{\sqrt{\mathrm{v}}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{5.65}$$

where the χ_s are the spin functions. The pair distribution function just describes the orbital part, so that one averages over spin functions. These spin averages are $\langle\chi_\uparrow\chi_\uparrow\rangle = 1$, $\langle\chi_\uparrow\chi_\downarrow\rangle = 0$, etc. The two pair distribution functions are

$$\begin{aligned} g_{\uparrow\downarrow}(\mathbf{r}_1 - \mathbf{r}_2) &= \frac{1}{N(N-1)}\sum_{\mathbf{k}_1\mathbf{k}_2}(|e^{i\mathbf{k}_1\cdot\mathbf{r}_1+i\mathbf{k}_2\cdot\mathbf{r}_2}|^2 + |e^{i\mathbf{k}_1\cdot\mathbf{r}_2+i\mathbf{k}_2\cdot\mathbf{r}_1}|^2) \\ &= \frac{2}{N(N-1)}\left(\frac{N}{2}\right)^2 = \tfrac{1}{2} \end{aligned} \tag{5.66}$$

$$\begin{aligned} g_{\uparrow\uparrow}(\mathbf{r}_1 - \mathbf{r}_2) &= \frac{1}{N(N-1)}\sum_{\mathbf{k}_1\mathbf{k}_2}|e^{i\mathbf{k}_1\cdot\mathbf{r}_1+i\mathbf{k}_2\cdot\mathbf{r}_2} - e^{i\mathbf{k}_1\cdot\mathbf{r}_2+i\mathbf{k}_2\cdot\mathbf{r}_1}|^2 \\ &= \frac{2}{N(N-1)}\sum_{\mathbf{k}_1\mathbf{k}_2}[1 - e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot(\mathbf{r}_1-\mathbf{r}_2)}] \\ &= \tfrac{1}{2}[1 - \Lambda(\mathbf{r}_1 - \mathbf{r}_2)^2] \end{aligned} \tag{5.67}$$

$$\Lambda(\mathbf{r}) = \frac{2}{N}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{5.68}$$

The summation over $\mathbf{k}$ runs only over occupied states. The antiparallel spin distribution function $g_{\uparrow\downarrow} = g_{\downarrow\uparrow} = \frac{1}{2}$ in the Hartree–Fock approximation. The cross term which results from the determinant is zero because of the orthogonality of the two spin functions. There is no correlation in the position of electrons of opposite spin. The parallel spin $g_{\uparrow\uparrow} = g_{\downarrow\downarrow}$ has a definite spatial dependence, which comes from the cross term which is retained since the spins are parallel and the spin averages are unity. The function $\Lambda(r)$ is calculated to be

$$\begin{aligned} \Lambda(r) &= \frac{2}{n_0}\int\frac{d^3k}{(2\pi)^3}n_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{3}{rk_F^3}\int_0^{k_F} kdk\sin(kr) \\ &= \frac{3}{rk_F}j_1(rk_F) \end{aligned} \tag{5.69}$$

where $j_1(x)$ is a spherical Bessel function. The two pair distribution functions are shown as the solid lines in Fig. 5.5. The parallel spin $g_{\uparrow\uparrow}(r)$ vanishes at $r = 0$ and approaches $\frac{1}{2}$ at large distances. The pair distribution function $g_{\uparrow\uparrow}(r)$ must vanish at $r = 0$ as a consequence of the exclusion principle for fermions. This feature is built into the many-particle wave function (5.62): if any pair of orbitals have the same spin and position, these two rows of the determinant are identical, which makes the determinant vanish.

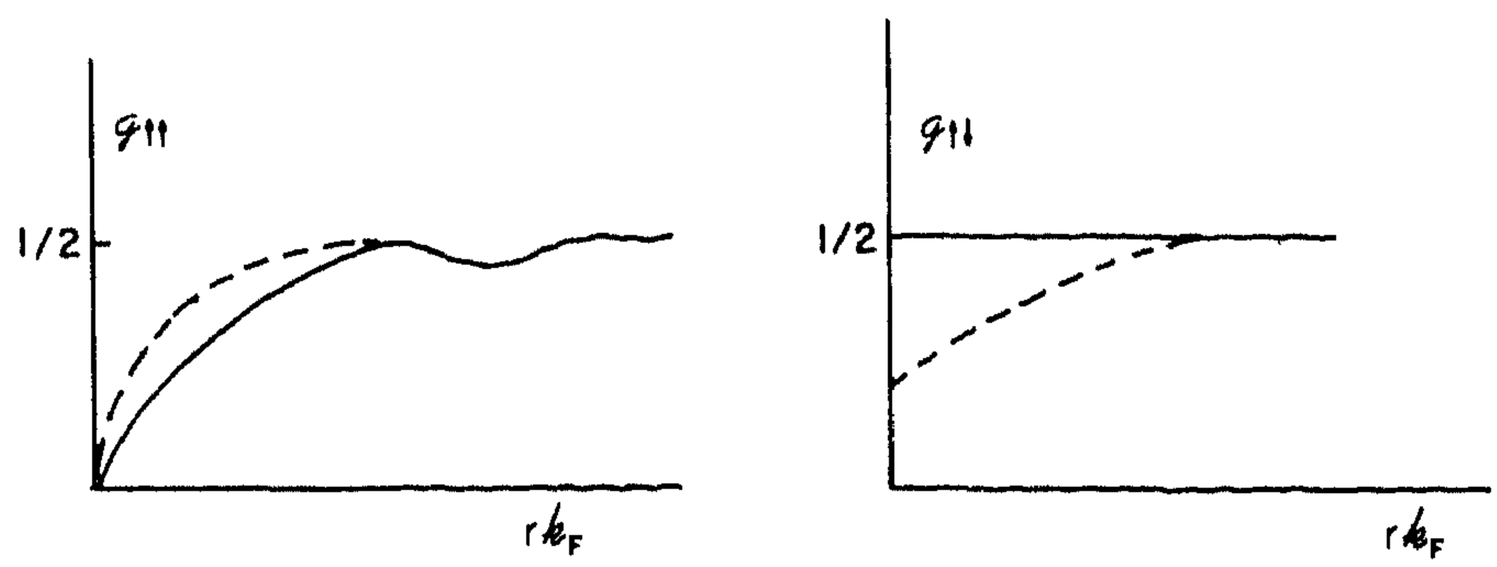


FIGURE 5.5 The pair distribution functions for parallel (left) and antiparallel (right) spins. The solid liine is the Hartree–Fock approximation, while the dashed line includes correlation.

The total pair distribution function for a particle is the combination of the result for parallel and antiparallel spin distributions:

$$g(r) = g_{\uparrow\uparrow}(r) + g_{\uparrow\downarrow}(r) \tag{5.70}$$

In a nonmagnetic system, the same result is obtained if the central particle has spin down. The pair distribution function has two features which are important for our discussion. First, there is the normalization integral:

$$n_0 \int d^3r[g(r) - 1] = -1 \tag{5.71}$$

The second is a statement about the ground state energy of the electron gas. It has kinetic energy plus Coulomb parts. The Coulomb energy around the hole in the Hartree–Fock approximation is

$$E_{\text{Coul}} = \frac{e^2 n_0}{2} \int d^3r \frac{g(r) - 1}{r} \tag{5.72}$$

The integral E_{Coul} is related to the exchange energy which is discussed in Sec. 5.6.3. The above two relations for $g(r)$ can be checked in the Hartree–Fock approximation. In this case $g(r) = \frac{1}{2} + g_{\uparrow\uparrow}(r)$. The normalization integral yields ($x = rk_F$)

$$n_0 \int d^3r[g_{\uparrow\uparrow}(r) - \tfrac{1}{2}] = -\frac{n_0}{2} \int d^3r \Lambda(r)^2 = -\frac{6}{\pi} \int_0^\infty dx j_1(x)^2 = -1$$

while the Coulomb integral yields

$$\begin{aligned} E_{\text{Coul}} &= \frac{e^2 n_0}{2} \int \frac{d^3r}{r} [g_{\uparrow\uparrow}(r) - \tfrac{1}{2}] = -\frac{e^2 n_0}{4} \int \frac{d^3r}{r} \Lambda(r)^2 \\ &= -\frac{3}{\pi} e^2 k_F \int_0^\infty dx \frac{j_1(x)^2}{x} = -\frac{3e^2 k_F}{4\pi} \end{aligned} \tag{5.73}$$

These results are correct for Hartree–Fock. It is the same exchange energy per particle which was derived by diagrammatic means.

The pair distribution function $g(r)$ is the distribution of electrons, on the average, about any electron. It goes to unity at large distances, which is a result of the uniform distribution n_0 of electron charge. That is, the electron charge density is $-en_0 g(r)$. Since $g(r)$ is less than unity near $r \to 0$, the electron charge is depleted in the vicinity of the electron. This reduction may be viewed as a hole in the electron density. Wigner suggested the name *exchange and correlation hole*. The hole moves with the electron.

According to the normalization (5.71), the total charge missing from this hole is one electron charge. The factor $-n_0(g - 1)$ is the density of hole charge, which integrates to 1. The sum rule is really a statement of charge neutrality. The hole has a positive charge density, since it is the absence of electrons. The integral of this charge density must be e. Each electron of charge $-e$ has in its hole the amount of charge e, and the system is neutral. The homogeneous electron gas has a uniform charge density only on the average. To a particular electron, the system is not uniform at all, since other electrons are not as likely to venture near to it as they are to other points.

The Coulomb integral gives the potential energy of the system. The electron interacts with its own hole charge. It is not influenced by the electrons or uniform charge density farther away, since they cancel on the average. Near the electron the positive background charge is not canceled by the other electrons, since the other electrons are not as likely to be

nearby. In the Hartree–Fock approximation, only electrons of parallel spin make the exchange and correlation hole. Since $g_{\uparrow\downarrow} = \frac{1}{2}$, the antiparallel spins are not affected.

It is possible to arrange $g(r)$ so that the sum rule (5.71) is still obeyed, yet the potential energy is lower (i.e., a larger negative value) than found in Hartree–Fock. This lowering is done by permitting $g_{\uparrow\downarrow}(r)$ to become less than $\frac{1}{2}$ near the point $r \to 0$, as shown by the dashed line in Fig. 5.5. There is no particular reason it must go to zero. The only fixed rule is that it may not be negative, since it is a probability. Since the total hole charge must be unity, the change in $g_{\uparrow\downarrow}(r)$ must be accompanied by a change in $g_{\uparrow\uparrow}(r)$ which causes it to rise more steeply with r, as shown by the dashed line in Fig. 5.5. It is these changes which give rise to the correlation energy. Changes in $g_{\uparrow\downarrow}$ and $g_{\uparrow\uparrow}$ will usually cost some kinetic energy, so that one cannot just adjust $g_{ss'}$ to maximize the potential energy alone. The system will seek the lowest energy state at $T = 0$, which is the state with some "correlation" between the motion of electrons with antiparallel spins. This "correlation" between the motion of pairs of electrons is what gives rise to the "correlation energy."

5.2. WIGNER LATTICE

Actual electron systems exist over a range of densities. The range in metals is approximately $1.8 < r_s < 6$. Electron densities inside atoms are also quite variable. It is necessary to have a description of the correlation energy which is valid for a distribution of densities. In the previous section, a formula was derived for the high-density limit of a homogeneous electron gas. The next step is to derive the result for the low-density limit. The final formulas are found by interpolating between these extremes. This procedure was first carried through by Wigner (1934), and his formula[1]

$$E_c = \frac{0.88}{r_s + 7.8} E_{ry} \tag{5.74}$$

is still widely used. Recent interpolation formulas are only slightly better. At low densities, Wigner speculated that the electrons would become localized and form a regular lattice. He used the jellium model where the positive charge is uniformly spread through the system. The electron lattice would presumably be a close-packed structure such as bcc, fcc, or hcp, in which electrons would vibrate around their equilibrium positions. There would be vibrational modes of the electrons, and they would be at the plasmon frequency. At very low density the potential energy becomes more important than the kinetic energy. The kinetic energy is in the zero-point motion of the vibrational modes. Localization cannot occur until the zero point energy is less than the potential energy.

Wigner calculated the potential energy in the following fashion. A Wigner–Seitz model was taken for the unit cell of the lattice. It is a sphere of radius $r_s a_0$, with the electron at the center. Each sphere has overall neutrality, since the one-electron charge at the center is canceled by the positive charge inside the volume of the sphere. For a sphere of radius $r_s a_0$, there is one unit of charge, which follows from the definition of r_s. Outside of each sphere, the electric field is zero. If all unit cells are neutral spheres, then they exert no electric fields on each other. The electric fields inside a sphere arise only from the electron and positive charge within that sphere. Of course this model is only approximate, since the unit cells are not truly spheres. Spheres cannot be packed together to cover all volume. The error made by the Wigner–Seitz approximation is remarkably small.

[1] Wigner's original paper contained an error. The corrected result is given.

The first term is the potential energy between the electron and the uniform positive background. It is

$$E_{ep} = \int d^3r\left(-\frac{e^2}{r}\right)n_0 = -\frac{3e^2}{r_s^3 a_0^3}\int_0^{r_s a_0} r dr$$
$$= -\frac{3e^2}{2a_0 r_s} = -\frac{3}{r_s}\left(\frac{e^2}{2a_0}\right) \tag{5.75}$$

where $-e^2/r$ is the potential energy and $n_0 = 3/(4\pi r_s^3 a_0^3)$ is the density of positive charge at each distance r. The result is $-3/r_s$ in atomic units, which is a large energy term.

The second term in the potential energy is the interaction of the positive charge with itself. The potential energy $V(r)$ from the positive charge at a distance r from the center is obtained by solving first for its equivalent electric field,

$$-\frac{dV}{dr} = eE(r) = \frac{e^2}{r^2}n_0\left(\frac{4\pi r^3}{3}\right) = \frac{e^2 r}{r_s^3 a_0^3} \tag{5.76}$$

which is e^2/r^2 times the total charge within the sphere of radius r. Integrating this equation to obtain $V(r)$, there is a constant of integration. It is determined by the condition that the total potential, from electron and positive charge, must vanish at the surface of the sphere. The constant is chosen to make $V(r)$ be $e^2/r_s a_0$ at the surface. The result is

$$V(r) = \frac{e^2}{2r_s a_0}\left[3 - \left(\frac{r}{r_s a_0}\right)^2\right] = \frac{1}{r_s}\left[3 - \left(\frac{r}{r_s a_0}\right)^2\right]E_{ry} \tag{5.77}$$

The potential energy of the positive charge interacting with itself is found using

$$E_{pp} = \frac{1}{2}\int d^3r V(r) n_0 = \frac{3}{4}\frac{e^2}{(r_s a_0)^4}\int_0^{r_s a_0} r^2 dr\left[3 - \left(\frac{r}{r_s a_0}\right)^2\right]$$
$$= \frac{3}{5}\frac{e^2}{r_s a_0} = \frac{6}{5r_s}E_{ry} \tag{5.78}$$

The result is multiplied by $\frac{1}{2}$ because it is a self-energy. The final result is $1.2/r_s$ in atomic units. There are only two potential energy terms. The interaction of the electron with itself is not included. Aside from the fact that it is infinity, it does not change in the metal and so does not contribute to the cohesive energy of the system.

The sum of these two terms is $-1.8/r_s$. It is the total potential energy in the Wigner lattice in the Wigner–Seitz approximation. It is a large term, e.g., it has a larger coefficient than the exchange and correlation energies which were found for the free-particle system. The system apparently has gained energy by the localization of the electrons.

The actual energy of several Wigner lattices was calculated by Sholl (1967) using Madelung summation methods. He found the energy of a lattice of point charges in a uniform positive background. He expressed his results as $-A/r_s$, where the parameter A was found for several lattices:

Lattice	A
sc	1.760
fcc	1.79175
bcc	1.79186
hcp	1.79168

The sc lattice has a unit cell which is not very spherical, and its A is different. But the other lattices are more close-packed, and the coefficient 1.792 is remarkably close to the Wigner–Seitz value of 1.8. The latter approximation errs by less than $\frac{1}{2}$%.

Is the Wigner lattice stable? To answer that question, first consider the potential energy on an individual electron in the vicinity of its equilibrium point. In the Wigner–Seitz model, it is just the negative of the potential energy from the positive charge:

$$V_e = -V(r) = -\frac{1}{r_s}\left[3 - \left(\frac{r}{r_s a_0}\right)^2\right]E_{ry} \tag{5.79}$$

At $r = 0$ it equals the prior result $-3/r_s$. Away from $r = 0$ the potential increases quadratically. If each electron moved independently of the others, then the electron would have harmonic vibrations about the equilibrium point. The zero-point motion of these vibrations is another term in the ground state energy, and includes the kinetic energy as well as some potential energy. The zero point energy was evaluated by Wigner, who found that its contribution toward the ground state energy was proportional to $r_s^{-3/2}$. At large r_s this term becomes smaller than the potential energy term, which falls off as r_s^{-1}. He concluded the lattice was stable at sufficiently large values of r_s. Of course the particles do not move independently. Their movements form collective vibrational modes. A better test of lattice stability is to calculate all the phonon modes and show that their frequencies are all real so the lattice is stable for collective motions. This calculation was done by Carr (1961). A summary of the theory of the Wigner lattice was given by Care and March (1975).

In the limit of low density, as $r_s \to \infty$, the potential energy of the Wigner lattice is proportional to $-1.792/r_s$. This term is the total Coulomb energy, which is correlation plus exchange. If the exchange energy $-0.9163/r_s$ is subtracted from this result, there remains the low-density limit of the correlation energy:

$$\lim_{r_s \to \infty} E_c = -\frac{0.8757}{r_s} \tag{5.80}$$

The numerator is rounded off to 0.88. Wigner also estimated that the high-density limit of the correlation energy was

$$\lim_{r_s \to 0} E_c = -0.113 \tag{5.81}$$

His estimate is now known to be incorrect, because of the $\ln r_s$ term which causes a divergence in E_c at small values of r_s. The final formula (5.74) is a simple interpolation scheme which satisfies these two limits. This formula is plotted in Fig. 5.6 as the line marked W. The interpolation procedure assumes that the correlation energy is a smooth function between the limits of high and low density. The interpolation would be reasonable if the system were in the same state in both limits. But the two limits describe systems in different phases. The high-density system has free-electron states, and the low-density system has localized electrons. The smooth interpolation through a phase boundary is not obviously correct, nor incorrect.

Other interpolation schemes have been proposed. One by Nozieres and Pines (1958) is

$$E_c = -0.115 + 0.031 \ln(r_s) \tag{5.82}$$

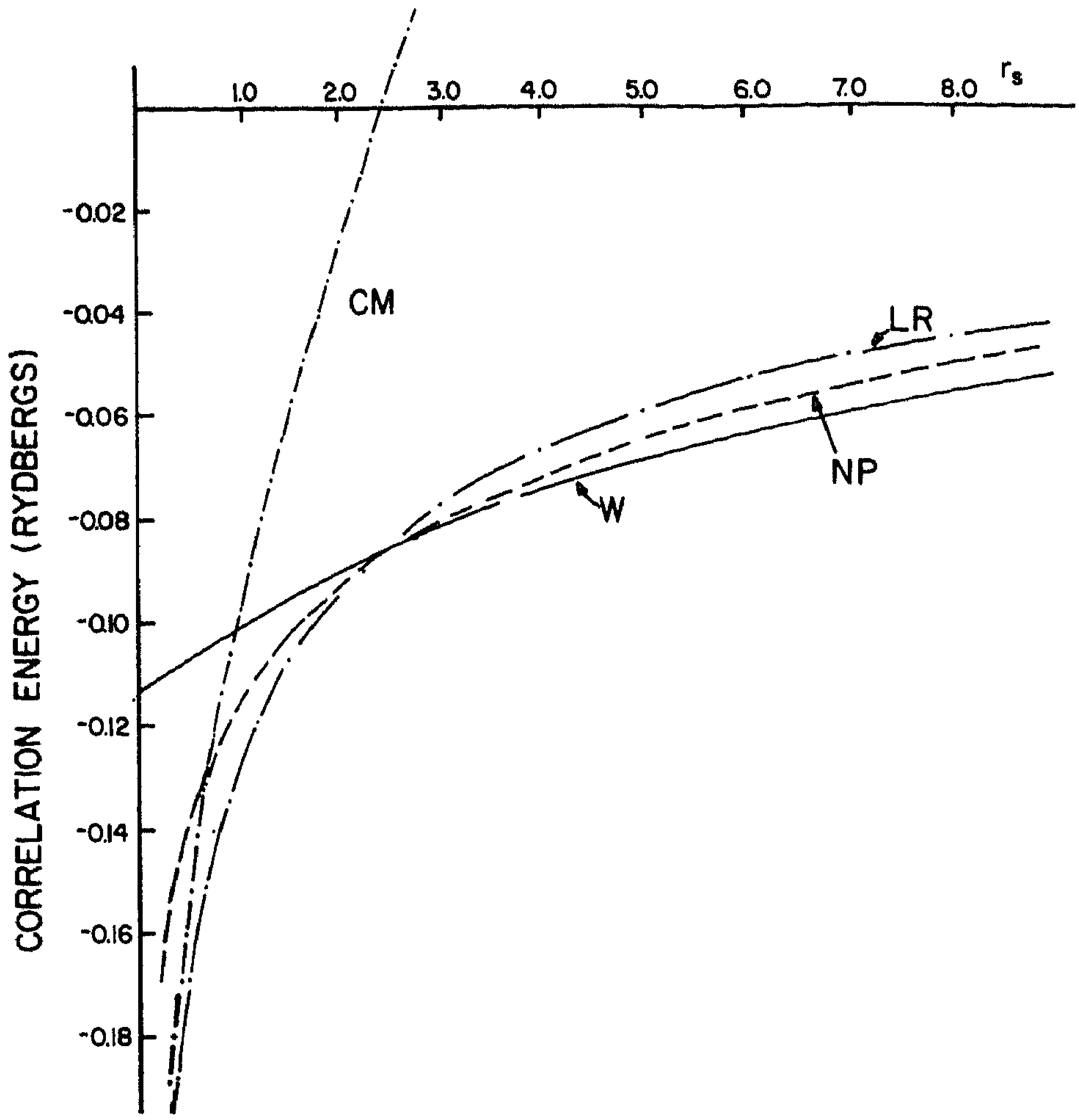


FIGURE 5.6 The correlation energy of the electron gas. The line CM (Carr–Maradudin) is the same small r_s expansion given in Fig. 5.4. The other curves are interpolation schemes according to Wigner (W), Nozieres and Pines (NP), and Lindgren and Rosen (LR).

It is shown in Fig. 5.6 as the curve marked NP. It is quite similar to Wigner's result in the range of metallic densities. Another interpolation formula has been proposed by Lindgren and Rosen (1970) (LR),

$$E_c = -\left[r_s + 3 + 4\sqrt{r_s} - \frac{0.08}{\sqrt{r_s}}\right]^{-1} \tag{5.83}$$

which is also shown in Fig. 5.6. It was chosen to agree with Carr and Maradudin at low values of r_s. Although the LR formula does not have a $\ln r_s$ term, it does agree well with the exact high-density expansion at low values of r_s. In the range of metallic densities, it has a lower (in magnitude) correlation energy than the previous two formulas.

There have been other calculations of the correlation energies at metallic densities. These are usually based on (5.56). The quantity $\varepsilon_{\mathrm{RPA}}(q, i\omega)$ is replaced by a better dielectric function—one which is more accurate at low densities. This approximation leads to an improved correlation energy. Even a direct evaluation of (5.56) gives a reasonable correlation energy for metallic densities (Lundqvist, 1969). There is no need to evaluate only the low r_s expansion of this result, as we did before. One can numerically solve it for all values of r_s. Some of these improved dielectric functions are discussed in Sec. 5.5.

5.3. METALLIC HYDROGEN

It has not been possible to study the three-dimensional Wigner lattice in a laboratory. There is another system, which is superficially quite similar, for which one could consider attempting experiments. It is metallic hydrogen. If one takes the Wigner lattice and changes the sign of all the charges, one has the model for metallic hydrogen. The protons are well approximated as point charges which are embedded in a free-electron gas. Since all other monovalent atoms form metals, perhaps hydrogen may also be made metallic.

Of course, at ordinary pressures, hydrogen does not form a metal. It is the gas H_2 at room temperatures. The lowering of temperature turns it into a liquid and then a solid. In both condensed phases the molecule H_2 retains its identity, so that the solid is a molecular crystal and an insulator. It may be possible to make hydrogen a metal by the application of pressure. As modern techniques achieve higher laboratory pressures, new phases of solid hydrogen continue to be discovered. However, they all seem to be molecular crystals and insulators (Mao and Hemley, 1994).

The discussion of metallic hydrogen will adopt the approximation that the electrons are a uniform electron gas. Better calculations show that the electron density is slightly nonuniform, since the electron density is higher near the proton than at the edge of the unit cell. Nevertheless, the assumption of uniformity makes only a small error in the analysis of the ground state energy. For a homogeneous electron gas, the electronic contribution to the ground state energy is the familiar result

$$E_g = \frac{2.2099}{r_s^2} - \frac{0.9163}{r_s} + E_c(r_s) + E_H \tag{5.84}$$

where E_H is the Hartree energy, and $E_c(r_s)$ is one of the formulas for the correlation energy. The first three terms for the ground state energy are for the case where the positive charge is spread uniformly throughout the system. In metallic hydrogen, the positive charge is a regular lattice of points. The extra Coulomb energy which results from the localization of the positive charge was calculated above for the Wigner lattice. Changing the sign of all the charges yields the same result: $E_H = -1.792/r_s$. The Hartree energy is the Coulomb energy calculated in the Hartree approximation, which neglects exchange and correlation.

There are, however, some subtle aspects to this calculation worth mentioning. The interaction energy between the protons and the uniform electron gas is calculated as before. It includes the interactions among the protons, but omits the proton interacting with itself. The electron–electron interactions are evaluated assuming a constant charge density n_0. Consider the Wigner–Seitz model of a spherical unit cell. Although there is one unit of electron charge in each unit cell, it is not treated as a single electron, since the electron–electron interaction would then be an electron interacting with itself. Instead, each electron is spread uniformly throughout the solid. In the free-electron model, the unit cell has one charge because 10^{23} electrons each contribute 10^{-23} of their charge. Then the calculation of electron–electron interactions has a negligible contribution from one electron interacting with itself. This picture is changed by the concept of the exchange and correlation hole. The electrons are not charges which are uniformly spread throughout the material. They are points which, on the average, may be found with uniform probability anywhere. But if one electron is at a point, the others are not, on the average, within its exchange and correlation hole. The radius of the exchange-correlation hole is similar to the radius of the Wigner–Seitz cell. The influence of the correlation hole is included in the correlation energy. Once it is included, the lattice energy from the protons is found by assuming the electron density is uniform.

The net result is that the energy from electron–electron interactions is small. In the Wigner–Seitz spherical model, the total energy from electron–electron interactions is

$$E_{ee} = E_x + E_c + \frac{1.2}{r_s} = \frac{0.237}{r_s} + E_c(r_s) \tag{5.85}$$

E_{ee} is the sum of exchange, correlation, and the term $1.2/r_s$ which comes from the uniform density of electrons interacting with themselves in the spherical unit cell. The term $0.28/r_s$ is small and positive, while E_c is small and negative. They largely cancel, so the net electron–electron energy is small. The concept of the exchange-correlation hole is correct. If one electron is put into a unit cell, then other electrons will not likely be there because of this hole. And if other electrons are not in the cell, there are no electron–electron interactions. Thus it is easy to understand why the net electron–electron interaction is small.

The ground state energy for metallic hydrogen is

$$\begin{aligned} E_g &= \frac{2.2099}{r_s^2} - \frac{1}{r_s}(0.916 + 1.792) + E_c(r_s) \\ &= \frac{2.210}{r_s^2} - \frac{2.708}{r_s} + E_c(r_s) \end{aligned} \tag{5.86}$$

$E_g(r_s)$ must be minimized, with respect to r_s, to obtain the predicted density of metallic hydrogen; the predicted value of r_s is defined as r_{s0}. The minimization will be done in the Hartree–Fock approximation by ignoring the correlation energy E_c, because the variation of $E_c(r_s)$ with respect to r_s is small at metallic densities. The correlation energy will have little influence upon the choice of r_{s0}. The simple equation to minimize is:

$$0 = \frac{dE_{HF}}{dr_s} = \frac{d}{dr_s}\left(\frac{2.210}{r_s^2} - \frac{2.708}{r_s}\right)_{r_{s0}} = -\frac{4.420}{r_{s0}^3} + \frac{2.708}{r_{s0}^2} \tag{5.87}$$

The minimum occurs at $r_{s0} = 1.632$, and the Hartree–Fock energy at this value is

$$E_{HF}(r_{s0}) = \frac{2.210}{r_{s0}^2} - \frac{2.708}{r_{s0}} = -0.830E_{ry} \tag{5.88}$$

The Wigner correlation energy at this value of r_{s0} is -0.093, so that the predicted ground state energy is $-0.923E_{ry}$. It is the binding energy per electron. It is not even as large as atomic hydrogen, which is 1.00 Rydberg per electron. The prediction is that metallic hydrogen is not as bound as atomic hydrogen. Of course, it is possible to do a better calculation of E_g for metallic hydrogen, which includes the nonuniformity of the electron gas. The best result so far is $E_g = -1.048E_{ry}$ per atom at $r_{s0} = 1.60$ by Hammerberg and Ashcroft (1974). Although this value is more bound than atomic hydrogen, it is still not as bound as molecular hydrogen or solid molecular hydrogen which has $r_s = 3.3$. The conclusion is that hydrogen would rather be a molecule than a metal, and this conclusion is in accord with experiments.

5.4. LINEAR SCREENING

Screening is one of the most important concepts in many-body theory. Charges, which are able, will move in response to an electric field. This charge movement will stabilize into a new distribution of charge around the electric field. This new distribution is just the right

amount of charge to cancel the electric field at large distances. The proof of this is rather trivial. If the electric field is not canceled at large distances, more charge will still be attracted until it is sufficient for cancellation. If the electric field is caused by an impurity charge distribution $\rho_i(\mathbf{r})$, with net charge $Q_i = \int d^3r\rho_i(r)$, the amount of mobile charge attracted to the surroundings is exactly $-Q_i$. This fact was already used in discussing the Friedel sum rule in Sec. 4.1. The name screening charge is applied to the mobile charge attracted by the impurity electric field. It will also have its own distribution in space $\rho_s(\mathbf{r})$. The screened potential from the impurity charge and the screening charge is given by

$$\phi(\mathbf{r}) = \int d^3r' \frac{\rho_i(\mathbf{r}') + \rho_s(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{5.89}$$

This formula is an exact result, as long as $\rho_s(\mathbf{r})$ is found exactly.

The screening charge is not necessarily in bound states due to the electric field from the impurity. Of course, that could happen if the electric field from the impurity is strong enough. But quite often the screening charge is from the unbound conduction electrons of the metal or semiconductor. In their motion through the crystal, they spend a little more time near the impurity potential, if it is attractive, than they do elsewhere in the solid. When these motions are averaged, there is more electron density near the impurity than elsewhere, which is the screening charge. If the impurity potential is repulsive for electrons, they tend to spend less time near the impurity, so the average charge is depleted near the impurity. Here the screening charge is positive, since it signifies a reduction in the average density of electrons, which have negative charge.

The classical macroscopic theory is quite familiar. The electric field $\mathbf{E}$ and displacement field $\mathbf{D}$ obey the equations

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = 4\pi\rho_i(\mathbf{r}) \tag{5.90}$$

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = 4\pi[\rho_i(\mathbf{r}) + \rho_s(\mathbf{r})] \tag{5.91}$$

All equations are Fourier-transformed to give

$$i\mathbf{q} \cdot \mathbf{D}(\mathbf{q}) = 4\pi\rho_i(\mathbf{q}) \tag{5.92}$$

$$i\mathbf{q} \cdot \mathbf{E}(\mathbf{q}) = 4\pi[\rho_i(\mathbf{q}) + \rho_s(\mathbf{q})] \tag{5.93}$$

The components of $\mathbf{D}(\mathbf{q})$ and $\mathbf{E}(\mathbf{q})$ along the direction $\mathbf{q}$ are the longitudinal fields $D_l(\mathbf{q})$ and $E_l(\mathbf{q})$. The longitudinal electric field is related to the scalar potential $E_l(\mathbf{r}) = -\nabla\phi(\mathbf{r})$ or its transform $\phi(\mathbf{q}) = iE_l(\mathbf{q})/q$:

$$D_l(\mathbf{q}) = \frac{4\pi}{iq}\rho_i(\mathbf{q}) \tag{5.94}$$

$$E_l(\mathbf{q}) = \frac{4\pi}{iq}[\rho_i(\mathbf{q}) + \rho_s(\mathbf{q})] \tag{5.95}$$

$$\phi(\mathbf{q}) = \frac{4\pi}{q^2}[\rho_i(\mathbf{q}) + \rho_s(\mathbf{q})] \tag{5.96}$$

The *dielectric response function* is defined as the ratio $D_l(\mathbf{q})/E_l(\mathbf{q})$ in the limit where $\rho_i \to 0$:

$$\varepsilon(\mathbf{q}) = \lim_{\rho_i \to 0} \frac{D_l(\mathbf{q})}{E_l(\mathbf{q})} = \lim_{\rho_i \to 0}\left[\frac{\rho_i(\mathbf{q})}{\rho_i(\mathbf{q}) + \rho_s(\mathbf{q})}\right] \tag{5.97}$$

In this limit $\varepsilon(\mathbf{q})$ becomes a property of the material and is independent of the charge distribution. One of our goals is to calculate this dielectric function. The linear screening model assumes this definition is true for nonzero $\rho_i(\mathbf{q})$, which gives for the potential

$$\phi'(\mathbf{q}) = \frac{4\pi}{q^2}\frac{\rho_i(\mathbf{q})}{\varepsilon(\mathbf{q})} \tag{5.98}$$

$$\phi'(\mathbf{r}) = \int \frac{d^3q}{(2\pi)^3}\frac{4\pi}{q^2}\frac{\rho_i(\mathbf{q})}{\varepsilon(\mathbf{q})}e^{i\mathbf{q}\cdot\mathbf{r}} \tag{5.99}$$

The potential ϕ' is the total potential from screening charge and impurity charge. That is obvious from its definition, since $\mathbf{E}(\mathbf{r})$ is the electric field in (5.90) from both screening and impurity charges. The potential ϕ' should be similar to the exact screened potential ϕ in (5.89). They are identical in the limit where ρ_i is small. The linear screening approximation is to calculate $\phi'(\mathbf{r})$ in place of $\phi(\mathbf{r})$. It is a much easier function to calculate, at least after $\varepsilon(\mathbf{q})$ has been determined. Another feature of the linear screening model is that the screening charge $\rho_s(\mathbf{q})$ density is proportional, in $\mathbf{q}$ space, to the impurity charge density $\rho_i(\mathbf{q})$. Linear screening models assume that

$$\varepsilon(\mathbf{q}) = \frac{\rho_i(\mathbf{q})}{\rho_i(\mathbf{q}) + \rho_s(\mathbf{q})} \tag{5.100}$$

$$\frac{\rho_s(\mathbf{q})}{\rho_i(\mathbf{q})} = \frac{1}{\varepsilon(\mathbf{q})} - 1 \tag{5.101}$$

are valid for finite values of ρ_i, rather than infinitesimal ones.

A simple example will illustrate when linear screening does not apply. Consider a point charge $Z = -1$ in the homogeneous electron gas with high density ρ_0. Here the screening charge is positive, which requires a depletion of the electron density around the impurity. This behavior is sketched in Fig. 5.7(a). Assume that the linear screening model works well for

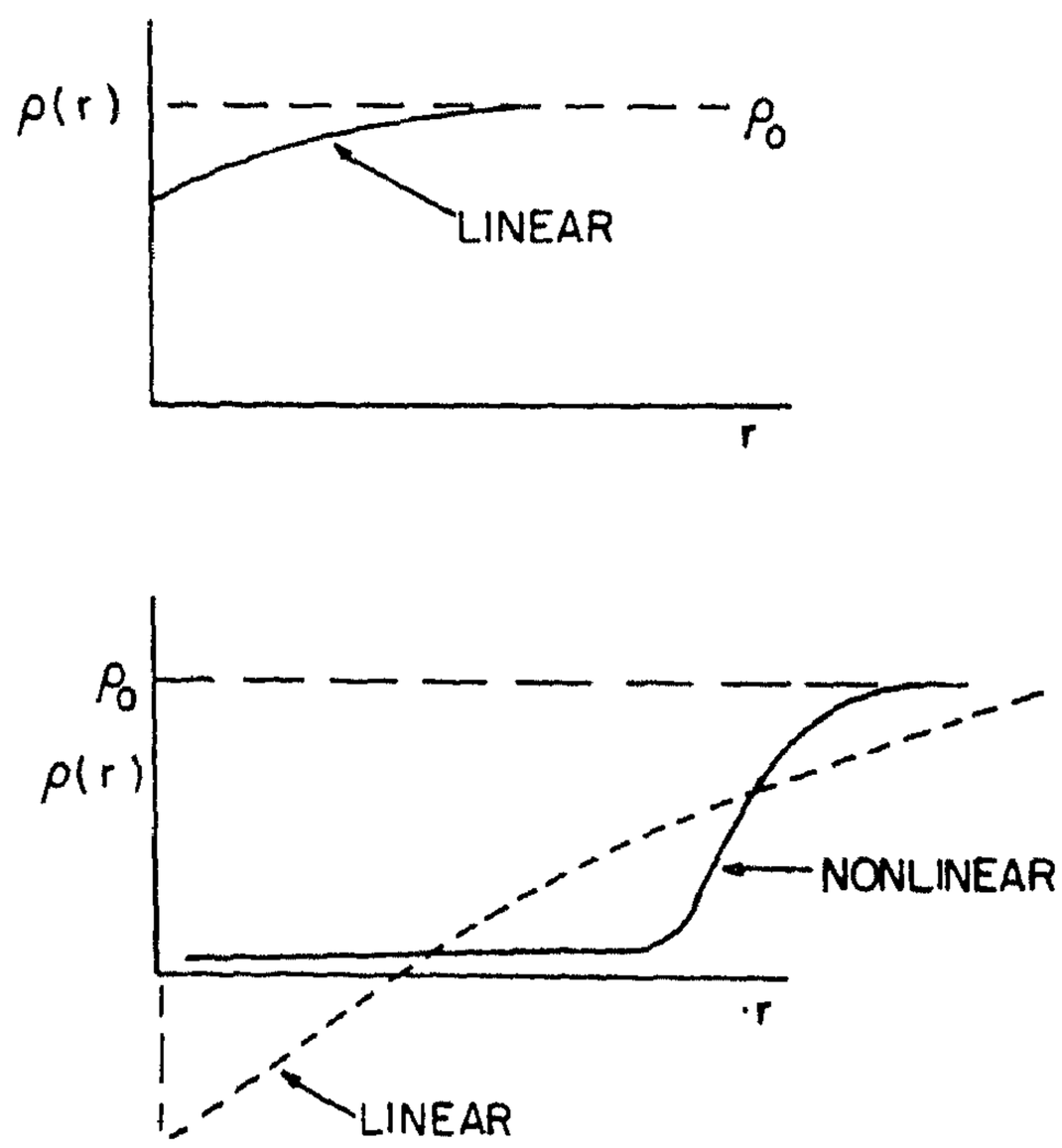


FIGURE 5.7 (Top) Change in electron charge density from linear screening (neglecting Friedel oscillations). (Bottom) A nonlinear screening model must be used for a much larger impurity charge.

$Z = -1$, although that may not be a good assumption in metals. Now consider the response to a large charge of $Z = -10$. In the linear response theory, the screening charge is multiplied 10-fold, so that it predicts the electron density $\rho(r)$ to have the dashed line in Fig. 5.7(b). Of course, this prediction is nonsense, since it has the electron particle density being negative for small values of r. The actual electron density has the form shown by the solid line. In this case the screening charge is not just 10 times that for $Z = -1$.

The macroscopic theory defined the dielectric function $\varepsilon(\mathbf{q})$. But it did not provide a clue to the technique of finding it. That is the role of a microscopic theory. The immediate goal is to derive a rigorous definition of $\varepsilon(\mathbf{q})$ in terms of microscopic operators. In the following section it is shown how this exact equation is solved, approximately, to give model dielectric functions which are used in calculations.

The derivation of the exact equation for $\varepsilon(\mathbf{q})$ starts by considering the interaction between two impurity charges $Z_1 e$ and $Z_2 e$ in the homogeneous electron gas. The interaction potential between these two charges, in the linear screening model, is proportional to the product $Z_1 Z_2$. The ground state energy of the system is evaluated, and all energy terms are extracted which are proportional to $Z_1 Z_2$. The summation of these terms is defined as the interaction potential between the two charges. Of course, there will also be terms proportional to Z_1^n or Z_2^n ($n = 2, 3, \ldots$), which are the energies needed to put each separate charge in by itself. The terms $Z_1^n Z_2^m$ for $m, n \geq 2$ are contributions to the nonlinear interaction. All other terms are ignored except $Z_1 Z_2$, since the immediate interest is the derivation of linear screening.

The Hamiltonian of the homogeneous electron gas, with two impurity charges $Z_1 e$ and $Z_2 e$ at $\mathbf{R}_1$ and $\mathbf{R}_2$, is written as

$$H = H_0 + \frac{Z_1 Z_2 e^2}{|\mathbf{R}_1 - \mathbf{R}_2|} - \frac{1}{\mathrm{v}} \sum_{\mathbf{q} \neq 0} v_q \rho(\mathbf{q}) \sum_{j=1}^{2} Z_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{5.102}$$

$$\rho(\mathbf{q}) = \sum_{\mathbf{p}\sigma} C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{5.103}$$

The first term H_0 is the Hamiltonian for the homogeneous electron gas (5.1). It includes the electron–electron interactions, as well as the kinetic energy. The term $e^2 Z_1 Z_2 / |\mathbf{R}_1 - \mathbf{R}_2|$ is the direct interaction between the two charges. The last term in (5.102) is the interaction potential between each impurity charge $Z_j e$ and the electrons of the homogeneous electron gas. They are represented by their density operator $\rho(\mathbf{q})$.

The ground state energy is calculated from the thermodynamic potential, which is found from the linked cluster theorems of (3.264)–(3.266):

$$\Omega = \Omega_0 - \frac{1}{\beta} \sum_{l=1}^{\infty} U_l$$

$$U_l = \frac{(-1)^l}{l} \int_0^{\beta} d\tau_1 \cdots \int_0^{\beta} d\tau_l \langle T_\tau \hat{V}(\tau_1) \cdots \hat{V}(\tau_l) \rangle_{\text{different connected}}$$

where the U_l are the different connected diagrams. Only the terms in this series will be evaluated which are proportional to $Z_1 Z_2$. The noninteracting thermodynamic potential Ω_0 comes from H_0. At zero temperature, Ω_0 is the ground state energy of the homogeneous electron gas, which was evaluated in Secs. 5.1 and 5.2. The last two terms in (5.102) are the interaction potential V which enters the perturbation expansion for the thermodynamic potential.

The term U_1 in the expansion has only one power of V. In this order, the only contribution proportional to Z_1Z_2 is from the direct interaction

$$U_1 = -\beta \frac{Z_1 Z_2 e^2}{|\mathbf{R}_1 - \mathbf{R}_2|} \equiv -\beta Z_1 Z_2 \int \frac{d^3q}{(2\pi)^3} v_q e^{i\mathbf{q}\cdot(\mathbf{R}_1-\mathbf{R}_2)} \tag{5.104}$$

The other first-order term is zero, since the average of $\rho(\mathbf{q})$ is zero in the electron gas unless $\mathbf{q} = 0$. Of course, these averages are taken with the Hamiltonian H_0, which is without the impurities present.

The term U_2 has two powers of V. Two powers of the last term in (5.102) give a contribution

$$\begin{aligned} U_2 = &\frac{1}{2\nu^2}\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \sum_{\mathbf{q}\mathbf{q}'} v_q v_{q'} \langle T_\tau \rho(\mathbf{q}, \tau_1)\rho(\mathbf{q}', \tau_2)\rangle \\ &\times (Z_1 e^{i\mathbf{q}\cdot\mathbf{R}_1} + Z_2 e^{i\mathbf{q}\cdot\mathbf{R}_2})(Z_1 e^{i\mathbf{q}'\cdot\mathbf{R}_1} + Z_2 e^{i\mathbf{q}'\cdot\mathbf{R}_2}) \end{aligned} \tag{5.105}$$

This term is the only one in U_2; the other perturbation $e^2Z_1Z_2/|\mathbf{R}_1 - \mathbf{R}_2|$ does not enter again since it has only the one connected diagram. In U_2 there is a term proportional to Z_1^2 and another proportional to Z_2^2. They are part of the energy needed to put each charge, separately, into the homogeneous electron gas. It is not the total energy, since there are terms in higher orders of perturbation theory, proportional to $Z_j^n (n \geq 3)$, which also contribute. The term U_2 also has a cross term, which is proportional to Z_1Z_2

$$U_2 = \frac{Z_1 Z_2}{\nu^2} \sum_{\mathbf{q}\mathbf{q}'} e^{i\mathbf{q}\cdot\mathbf{R}_1 + i\mathbf{q}'\cdot\mathbf{R}_2} v_q v_{q'} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \langle T_\tau \rho(\mathbf{q}, \tau_1)\rho(\mathbf{q}', \tau_2)\rangle$$

There are no other terms proportional to Z_1Z_2. The higher linked cluster terms U_l have only higher powers of the charges. The derivation is completed. The net interaction between the two charges is the sum of the terms from U_1 and U_2. The U_2 term is simplified by the fact that in a homogeneous system it is nonzero only when $\mathbf{q} + \mathbf{q}' = 0$, since the density–density correlation function of the electron operators is nonzero only in this circumstance:

$$\Delta\Omega = Z_1 Z_2 \int \frac{d^3q}{(2\pi)^3} v_q e^{i\mathbf{q}\cdot(\mathbf{R}_1-\mathbf{R}_2)} \tag{5.106}$$

$$\times \left[1 - \frac{v_q}{v\beta}\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \langle T_\tau \rho(\mathbf{q}, \tau_1)\rho(-\mathbf{q}, \tau_2)\rangle\right] \tag{5.107}$$

This formula is compared with the linear screening model (5.99) for the potential from a charge distribution. One charge, say Z_1, is the impurity ($\rho_i(\mathbf{q}) = Z_1$). The other charge Z_2 is the test charge which measures the strength of the screened potential. The net interaction between the charges can be written as a screened coulomb interaction of the form

$$V_s(\mathbf{R}_1 - \mathbf{R}_2) = \Delta\Omega = Z_1 Z_2 \int \frac{d^3q}{(2\pi)^3} \frac{v_q}{\varepsilon(\mathbf{q})} e^{i\mathbf{q}\cdot(\mathbf{R}_1-\mathbf{R}_2)} \tag{5.108}$$

which provides a rigorous definition of the dielectric function:

$$\frac{1}{\varepsilon(\mathbf{q})} = 1 - \frac{v_q}{\nu\beta}\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \langle T_\tau \rho(\mathbf{q}, \tau_1)\rho(-\mathbf{q}, \tau_2)\rangle \tag{5.109}$$

The correlation term can be further simplified. The correlation function depends only on the difference of the two arguments $\tau_1 - \tau_2$. This fact, along with the periodicity (3.17) of the argument, permits one of the τ integrals to be eliminated to give

$$\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \langle T_\tau \rho(\mathbf{q}, \tau_1)\rho(-\mathbf{q}, \tau_2)\rangle = \beta \int_0^\beta d\tau \langle T_\tau \rho(\mathbf{q}, \tau)\rho(-\mathbf{q}, 0)\rangle$$

so the inverse dielectric function is

$$\frac{1}{\varepsilon(\mathbf{q})} = 1 - \frac{v_q}{\nu} \int_0^\beta d\tau \langle T_\tau \rho(\mathbf{q}, \tau)\rho(-\mathbf{q}, 0)\rangle \tag{5.110}$$

The exact result (5.110) is very important. It relates the dielectric function to the density–density correlation function. The time variation of the operators $\rho(\mathbf{q}, \tau) = e^{\tau H_0}\rho(\mathbf{q})e^{-\tau H_0}$ is governed by H_0 which is the full Hamiltonian for the homogeneous electron gas but without the potential of the impurities.

The static density–density correlation function $\langle \rho(\mathbf{q})\rho(-\mathbf{q})\rangle$ is related to the static structure factor $S(\mathbf{q})$, which was defined in Sec. 1.6. For a system of N_e electrons

$$\frac{1}{N_e} \langle \rho(\mathbf{q})\rho(-\mathbf{q})\rangle = N_e \delta_{\mathbf{q}=0} + S(\mathbf{q}) \tag{5.111}$$

The pair distribution function $g(\mathbf{r})$ of the electron gas can be obtained from a knowledge of $S(\mathbf{q})$. The latter quantity is important in describing correlation in the electron gas. This observation raises the question of whether there is any relationship between $S(\mathbf{q})$ and $1/\varepsilon(\mathbf{q})$. In fact, yes. In order to show this relationship, generalize (5.110) to nonzero values of frequency:

$$\frac{1}{\varepsilon(\mathbf{q}, i\omega_n)} = 1 - \frac{v_q}{\nu} \int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau \rho(\mathbf{q}, \tau)\rho(-\mathbf{q}, 0)\rangle \tag{5.112}$$

This result is true, although it has not been proved here. Actually, it could be regarded as the definition of the longitudinal dielectric function $\varepsilon(\mathbf{q}, i\omega_n)$.

Equation (5.112) is used to prove the following important theorem:

$$N_e \delta_{\mathbf{q}=0} + S(\mathbf{q}) = -\frac{1}{n_0 v_q} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{1}{1 - e^{-\beta\omega}} \operatorname{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] \tag{5.113}$$

The function $\varepsilon(\mathbf{q}, \omega)$ is the retarded function obtained from $\varepsilon(\mathbf{q}, i\omega_n)$ by $i\omega_n \to \omega + i\delta$. The subscript "ret" will be omitted, since the retarded function is always intended when writing $\varepsilon(\mathbf{q}, \omega)$. The factor $n_0 = k_F^3/3\pi^2$ is the particle density, and $v_q = 4\pi e^2/q^2$.

To prove this theorem, again introduce the sets of states $|m\rangle$ and $|n\rangle$, which are exact eigenstates of the Hamiltonian. These states were used in Sec. 3.3 for proving several theorems. The Hamiltonian of the homogeneous electron gas has $H_0|n\rangle = E_n|n\rangle$. These complete sets of states are introduced into (5.112):

$$\frac{1}{\varepsilon(\mathbf{q}, i\omega_n)} = 1 - \frac{v_q}{\nu} \int_0^\beta d\tau \sum_{nm} e^{-\beta E_n} |\langle n|\rho(\mathbf{q})|m\rangle|^2 e^{\tau(i\omega + E_n - E_m)}$$

The τ-integral can be done and the retarded function is found in a formal manner:

$$\frac{1}{\varepsilon(\mathbf{q}, i\omega)} = 1 - \frac{v_q}{\nu}\sum_{nm} e^{-\beta E_n}|\langle n|\rho(\mathbf{q})|m\rangle|^2 \frac{e^{\beta(E_n - E_m)} - 1}{i\omega + E_n - E_m}$$
$$\frac{1}{\varepsilon(\mathbf{q}, \omega)} = 1 - \frac{v_q}{\nu}\sum_{nm} e^{-\beta E_n}|\langle n|\rho(\mathbf{q})|m\rangle|^2 \frac{e^{\beta(E_n - E_m)} - 1}{\omega + E_n - E_m + i\delta} \tag{5.114}$$

The quantity on the right is real except for the factor of $i\delta$. The imaginary part of this term just gives a delta function:

$$\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] = \frac{\pi v_q}{\nu}\sum_{nm} e^{-E_n}|\langle n|\rho(\mathbf{q})|m\rangle|^2[e^{\beta(E_n - E_m)} - 1]\delta(\omega + E_n - E_m)$$
$$\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] = -(1 - e^{-\beta\omega})\frac{\pi v_q}{\nu}\sum_{nm} e^{-\beta E_n}|\langle n|\rho(\mathbf{q})|m\rangle|^2\delta(\omega + E_n - E_m) \tag{5.115}$$

Divide by $(1 - e^{-\beta\omega})$ and integrate over all ω:

$$\int_{-\infty}^{\infty}\frac{d\omega}{\pi}\frac{1}{1 - e^{-\beta\omega}}\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] = -\frac{v_q}{\nu}\sum_{nm} e^{-\beta E_n}|\langle n|\rho(\mathbf{q})|m\rangle|^2$$
$$= -\frac{v_q}{\nu}\langle\rho(\mathbf{q})\rho(-\mathbf{q})\rangle \tag{5.116}$$

The term on the far right is the static density–density correlation function. It gives the static structure factor $S(\mathbf{q})$, as shown in (5.111). This completes the proof of (5.113). At zero temperature, the above formula becomes ($\mathbf{q} \neq 0$)

$$S(\mathbf{q}) = -\frac{1}{n_0 v_q}\int_0^{\infty}\frac{d\omega}{\pi}\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] \tag{5.117}$$

which is the way it is often written. The pair distribution function $g(\mathbf{r})$ or the static structure factor $S(\mathbf{q})$ is obtained from a knowledge of the frequency-dependent dielectric function $\varepsilon(\mathbf{q}, \omega)$. The latter formula is not dependent on any assumptions regarding linear screening. It is exact. It arises because both $S(\mathbf{q})$ and $\varepsilon(\mathbf{q}, \omega)$ are related to the density–density correlation function. The assumption of linear screening is merely using (5.99) to calculate the screened potential from the impurity charge distribution $\rho_i(\mathbf{q})$.

The density–density correlation function

$$-\int_0^{\beta} d\tau e^{i\omega_n\tau}\langle T_\tau\rho(\mathbf{q}, \tau)\rho(-\mathbf{q}, 0)\rangle \tag{5.118}$$

has the appearance of a Green's function in the Matsubara representation. There is an operator $\rho(-\mathbf{q}, 0)$ acting at $\tau = 0$ and its inverse at τ. Since the operator is the density, the Green's function provides the response of the system to a density fluctuation. To develop the analogy further, the function

$$S(\mathbf{q}, \omega) = -\frac{1}{n_0 v_q}\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] \tag{5.119}$$

is the spectral function of this operator, since it is proportional to the imaginary part of the retarded Green's functions associated with this correlation function. This observation is important, since the spectral functions provide direct physical information. For the electron or the phonon, peaks in their spectral functions are interpreted as excitations of these operators.

In a similar way, the peaks in $S(\mathbf{q}, \omega)$ are interpreted as longitudinal excitations of the electron gas. These are two-particle excitations, since the density operator itself contains two operators—one creation and one destruction. In fact, some of the excitations, such as plasmons, are collective excitations of many particles.

The density operator has boson properties, since it is the product of two fermion operators. So $S(\mathbf{q}, \omega)$ is a spectral function for boson operators. Consequently, it has many features in common with other spectral functions for bosons; i.e., compare (5.113) with the similar phonon result (3.136):

$$2N_{\mathbf{q}} + 1 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} n_B(\omega) B(\mathbf{q}, \omega)$$

Another feature of $S(\mathbf{q}, \omega)$ is that it is positive for $\omega > 0$ and negative for $\omega < 0$, with $S(\mathbf{q}, -\omega) = -S(\mathbf{q}, \omega)$. This identity may be shown directly from (5.115).

The dielectric function has been defined in terms of the interaction between two fixed impurity charges. The assumption has been that the impurity charges are classical objects. There remains the question of the effective interaction between two electrons, which are surely not classical objects. The present theory suggests that there is an additional factor in the effective interaction between two electrons, which is a vertex correction

$$W(q) = \frac{v_q}{\varepsilon(q)} \Gamma(q) \tag{5.120}$$

The main contribution to the vertex correction $\Gamma(q)$ arises from the ladder diagrams at the end points of the interaction. The method of evaluating this vertex correction is discussed in later sections of this chapter.

5.5. MODEL DIELECTRIC FUNCTIONS

No one has yet derived the exact dielectric function of the homogeneous electron gas. Instead, approximate solutions have been obtained to (5.110). Some of these have been very successful, perhaps because they are simple or perhaps because they are accurate. They acquire the name of their inventor(s). Four of them are described here, which is only a small subset of the vast number which are available.

5.5.1. Thomas–Fermi

The Thomas (1927)–Fermi (1928) theory is also called the Fermi–Thomas theory about half the time. The derivation begins with the exact equation for the screened potential energy from an impurity charge distribution $\rho_i(\mathbf{r})$,

$$\nabla^2 V(\mathbf{r}) = 4\pi e[\rho_i(\mathbf{r}) + \rho_s(\mathbf{r})] \tag{5.121}$$

where $\rho_s(\mathbf{r})$ is the screening charge. The symbol ρ is used for charge density, and $n(\mathbf{r})$ for particle density ($\rho = -en$), and $\rho(\mathbf{q})$ for the density operator. This equation can be obtained from (5.89) by the application of $-e\nabla^2$. The factor $-e$ arises because $V(\mathbf{r})$ is the potential energy for electrons. Now start to make approximations. In Thomas–Fermi theory, the

electron density $n(\mathbf{r})$ is represented locally as a free-particle system. Write the screening charge as the difference between $n(\mathbf{r})$ and the equilibrium charge density n_0,

$$\rho_s(\mathbf{r}) = -e[n(\mathbf{r}) - n_0] \tag{5.122}$$

where the electron charge is $-e$. For a free-particle system, the local density is $n(\mathbf{r}) = k_F^3(\mathbf{r})/3\pi^2$, where the Fermi wave vector is now a local quantity. It, in turn, is determined by the condition that the chemical potential μ is independent of position:

$$\frac{k_F^2(\mathbf{r})}{2m} = E_F(\mathbf{r}) = \mu - V(\mathbf{r}) \tag{5.123}$$

The physics is illustrated in Fig. 5.8. Assume the potential $V(\mathbf{r})$ is slowly varying in space. Draw a little box in each region of space as the shaded region and treat it as a gas of fermions. If the absolute Fermi level is at μ, then the effective Fermi level is reduced or raised by the value of the local potential V. If these approximations are collected, there results the equation

$$\nabla^2 V = 4\pi e\left\{\rho_i(\mathbf{r}) + en_0 - en_0\left[1 - \frac{V(\mathbf{r})}{E_F}\right]^{3/2}\right\} \tag{5.124}$$

For atoms, this approximate equation is solved exactly with $E_F = 0$ and $\rho_i = Z\delta^3(\mathbf{r})$ to give a good description of atomic potentials and charge distributions (Landau and Lifshitz, 1958). The assumption that $V(\mathbf{r})$ is slowly varying does not seem unduly restrictive. To get a linear screening model, and hence a dielectric function, a further assumption is needed. It is assumed that $V/E_F \ll 1$, so the root is expanded $(1 - V/E_F)^{3/2} \approx 1 - 3V/2E_F$ to obtain the equation

$$\nabla^2 V = 4\pi e\rho_i(\mathbf{r}) + \frac{6\pi e^2 n_0}{E_F} V(\mathbf{r}) \tag{5.125}$$

The term on the right, proportional to V, is now moved to the left. Its coefficient is defined as the square of the Thomas–Fermi screening wave vector q_{TF}

$$(\nabla^2 - q_{TF}^2)V(\mathbf{r}) = 4\pi e\rho_i(\mathbf{r}) \tag{5.126}$$

$$q_{TF}^2 = \frac{6\pi e^2 n_0}{E_F} \tag{5.127}$$

This equation may be solved in Fourier transform space to give

$$V(\mathbf{r}) = -4\pi e \int \frac{d^3q}{(2\pi)^3} \frac{\rho_i(\mathbf{q})}{q^2 + q_{TF}^2} e^{i\mathbf{q}\cdot\mathbf{r}} \tag{5.128}$$

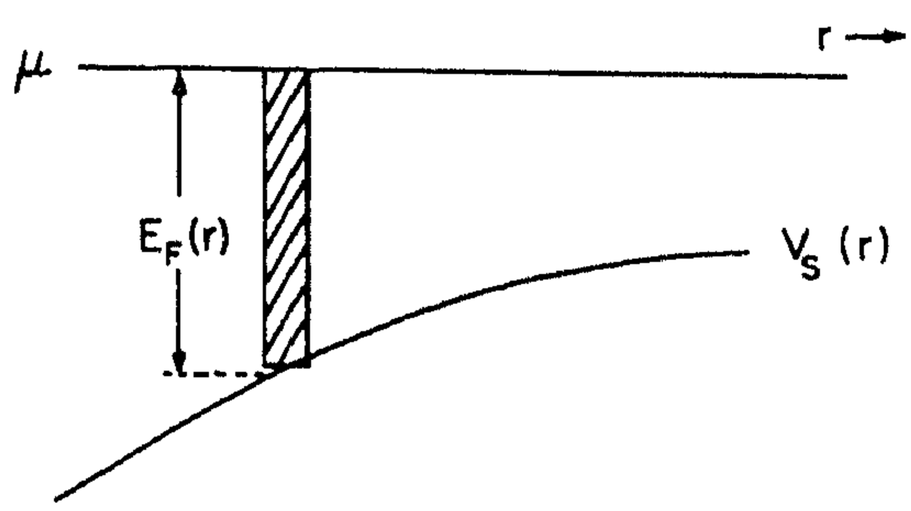


FIGURE 5.8 The Thomas–Fermi model has a Fermi degeneracy energy E_F which varies with position according to the variation in the potential energy.

Compare this equation with (5.98), and conclude that the Thomas–Fermi dielectric function is

$$\varepsilon(q) = 1 + \frac{q_{TF}^2}{q^2} \tag{5.129}$$

It has a simple form, which makes it easy to use in a variety of calculations. That is the chief explanation for its popularity.

For example, an analytical result can be obtained when the impurity is a point charge $\rho_i(\mathbf{q}) = Q_i$. The integrals to evaluate are ($\nu = \cos\theta$)

$$V(\mathbf{r}) = -\frac{eQ_i}{\pi}\int_0^\infty \frac{q^2 dq}{q^2 + q_{TF}^2}\int_{-1}^{1} d\nu e^{iqr\nu} \tag{5.130}$$

$$= -2\frac{eQ_i}{\pi r}\int_0^\infty \frac{q dq}{q^2 + q_{TF}^2}\sin(qr) \tag{5.131}$$

$$= -\frac{eQ_i}{i\pi r}\int_{-\infty}^{\infty} \frac{q dq e^{iqr}}{q^2 + q_{TF}^2} \tag{5.132}$$

$$V(\mathbf{r}) = -\frac{eQ_i}{r}e^{-q_{TF}r} \tag{5.133}$$

The last integral is done by closing the integration contour in the upper half plane and taking the residue of the pole at iq_{TF}. The screened interaction has the form of a Yukawa potential. The interaction declines rapidly at large distances because of the exponential dependence $\exp(-q_{TF}r)$. In metals, the Thomas–Fermi wave vector has a typical value of $1\,\text{Å}^{-1}$. The screened Coulomb potential declines rapidly on the scale of a unit cell. The screening wave vector may be expressed in atomic units as

$$a_0 q_{TF} = \left(\frac{4}{\pi}k_F a_0\right)^{1/2} = \frac{1.5632}{\sqrt{r_s}} \tag{5.134}$$

For example, at sodium density $r_s = 3.96$, one finds $q_{TF} = 1.48\,\text{Å}^{-1}$.

Thomas–Fermi theory provides only a static model for $\varepsilon(q)$. It is not usually used to describe the dynamic response $\varepsilon(\mathbf{q}, \omega)$.

5.5.2. Lindhard, or RPA

The Lindhard (1954) dielectric function is more commonly called the RPA, for random phase approximation. It is a model for a static $\varepsilon(q)$ or dynamic $\varepsilon(\mathbf{q}, \omega)$ dielectric function. It was already introduced in Sec. 5.1 for the discussion of correlation energies. It is rather easy to derive and has a simple conceptual basis. It also predicts correctly a number of properties of the electron gas such as plasmons. In the early days of electron gas theory, it was *the* dielectric function. Nowadays there is a tendency to use one of the recent models, which are better for describing the response of the electron gas. Two derivations of the RPA will be presented: one from equations of motion and the second using Green's functions and diagrams.

The derivation by equations of motion is also called the method of *self-consistent field* (Ehrenreich and Cohen, 1959). One introduces an impurity charge density $\rho_i(\mathbf{r}, t)$ or its equivalent Fourier transform $\rho_i(\mathbf{q}, \omega)$. The equivalent impurity potential is $V_i(\mathbf{q}, \omega) =$

$-e\phi_i(\mathbf{q}, \omega)$. There are three potential energies in the problem. The first is from the impurity $V_i(\mathbf{q}, \omega)$, the second is from the screening charge $V_s(\mathbf{q}, \omega)$, and the third is the total potential, which is the sum of these two:

$$V(\mathbf{q}, \omega) = V_i(\mathbf{q}, \omega) + V_s(\mathbf{q}, \omega) \tag{5.135}$$

$$\nabla^2 V_s(\mathbf{r}, t) = 4\pi e \rho_s(\mathbf{r}, t), \text{ or } V_s(\mathbf{q}, \omega) = -\frac{4\pi e}{q^2}\rho_s(\mathbf{q}, \omega) \tag{5.136}$$

$$\nabla^2 V_i(\mathbf{r}, t) = 4\pi e \rho_i(\mathbf{r}, t), \text{ or } V_i(\mathbf{q}, \omega) = -\frac{4\pi e}{q^2}\rho_i(\mathbf{q}, \omega) \tag{5.137}$$

The major assumption in the derivation is that the electrons respond to the total energy V. When solving the equations of motion for the electron, use the potential $V(\mathbf{r}, t)$ or its transform $V(\mathbf{q}, \omega)$. This choice presents a minor problem, since initially $V(\mathbf{q}, \omega)$ is not known. That is the object of the calculation, and $V_i(\mathbf{q}, \omega)$ is assumed known. Once $V(\mathbf{q}, \omega)$ is known, the dielectric function in the linear screening model is

$$\varepsilon(\mathbf{q}, \omega) = \frac{V_i(\mathbf{q}, \omega)}{V(\mathbf{q}, \omega)} \tag{5.138}$$

and the calculation is completed. In the method of self-consistent field, it is assumed that the electrons respond to V; then try to determine this function self-consistently. Write the effective Hamiltonian for the electrons as

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_p C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \frac{1}{\nu}\sum_{\mathbf{q}} V(\mathbf{q}, t)\rho(\mathbf{q}) \tag{5.139}$$

The time dependence $V(\mathbf{q}, t)$ is put directly into the Hamiltonian (5.139). The impurity charge is regarded as a classical system which is oscillating. The goal is to find the quantum response of the electron gas to this classical oscillation. Furthermore, the impurity is assumed to oscillate at a single frequency: $\rho_i(\mathbf{r}, t) = \rho_i(\mathbf{r})e^{-i\omega t}$ and $V_i(\mathbf{r}, t) = V_i(\mathbf{r})e^{-i\omega t}$. The average response of the system will depend on ω, so write the average of $\rho(\mathbf{q}, t)$ as $\langle\rho(\mathbf{q}, t)\rangle = \rho(\mathbf{q}, \omega)e^{-i\omega t}$ and the average of $V(\mathbf{q}, t)$ as $\langle V(\mathbf{q}, t)\rangle = V(\mathbf{q}, \omega)e^{-i\omega t}$. For the homogeneous electron gas, the density operator $\rho(\mathbf{q})$ has an expectation value of zero for $\mathbf{q} \neq 0$. When the impurity is present, the expectation value is nonzero and is proportional to the average for the screening charge:

$$\begin{aligned}\langle\rho_s(\mathbf{q}, t)\rangle &= -e\langle\rho(\mathbf{q}, t)\rangle = -e\sum_{\mathbf{p}\sigma}\langle C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}\rangle \\ &= -e\rho(\mathbf{q}, \omega)e^{-\omega t}\end{aligned} \tag{5.140}$$

Since the averages for ρ_s and ρ are proportional, it simplifies the discussion to use only one symbol, which we choose to be ρ. For example, in terms of the average $\langle\rho_s\rangle = -e\langle\rho\rangle$, Eq. (5.136) is

$$V_s(\mathbf{q}, \omega) = \frac{4\pi e^2}{q^2}\rho(\mathbf{q}, \omega) \tag{5.141}$$

The dielectric response function is defined as the ratio (5.138). The first term in (5.139) is the kinetic energy of the electrons, and the second is their interaction with the self-consistent potential $V(\mathbf{q}, t)$. Note that there are no explicit electron–electron interactions. They are included, indirectly, in the interaction term: the part of $V(\mathbf{q}, t)$ from the screening $V_s(\mathbf{q}, t)$ is caused by electron–electron interactions. It is a rather crude way to include these interactions,

since it includes screening but neglects all other effects of correlation and exchange. In later sections, other dielectric functions are introduced which are more accurate.

To obtain the screened potential V_s in (5.136), an expression is derived for the screening particle density $\rho_s(\mathbf{q}, t)$. It is obtained by writing an equation of motion for this operator and then solving it approximately. A perturbation on the system of $(\mathbf{q}, \omega)$ will cause polarization of the electron system, so that the average $\langle\rho(\mathbf{q}, t)\rangle$ will now have a nonzero value. In the linear screening model, this average is assumed to be proportional to the potential causing the perturbation, so $\langle\rho(\mathbf{q}, t)\rangle \propto \langle V(\mathbf{q}, t)\rangle$. The goal is to determine the constant of proportionality.

The equation of motion of the density operator comes from

$$i\frac{d}{dt}\rho(\mathbf{q}, t) = [H, \rho(\mathbf{q}, t)] \tag{5.142}$$

Actually, it is more convenient to evaluate the equation of motion for the operator:

$$i\frac{d}{dt}C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} = [H, C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma}] \tag{5.143}$$

After it has been evaluated, the result is summed over $(\mathbf{p}\sigma)$ to obtain $\rho(\mathbf{q}, t)$. The impurity potential $V(\mathbf{q}, t)$ is assumed to be oscillating at a single frequency $\exp(-i\omega t)$ so that the time derivative on the left of (5.102) gives $-i\omega C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma}$. The commutators on the right are evaluated for the Hamiltonian (5.139),

$$\sum_{\mathbf{k},5}\varepsilon_{\mathbf{k}}[C^\dagger_{\mathbf{k}s}C_{\mathbf{k}s}, C_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma}] = (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}})C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} \tag{5.144}$$

$$\begin{aligned}\frac{1}{\nu}\sum_{\mathbf{q}'\mathbf{k},s} V(\mathbf{q}', t)[C^\dagger_{\mathbf{k}+\mathbf{q}',s}C_{\mathbf{k}s}, C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma}]\\ = \frac{1}{\nu}\sum_{\mathbf{q}'} V(\mathbf{q}', t)(C^\dagger_{\mathbf{p}+\mathbf{q}+\mathbf{q}',\sigma}C_{\mathbf{P}\sigma} - C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}-\mathbf{q}',\sigma})\end{aligned} \tag{5.145}$$

which gives the equation

$$\begin{aligned}(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{q}+\mathbf{q}} + \omega)C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} &= \frac{1}{\nu}\sum_{\mathbf{q}'} V(\mathbf{q}', t)(C^\dagger_{\mathbf{p}+\mathbf{q}+\mathbf{q}',\sigma} - C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}-\mathbf{q}',\sigma})\\ &\approx \frac{1}{\nu}V(\mathbf{q}, t)(C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} - C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}+\mathbf{q},\sigma})\end{aligned} \tag{5.146}$$

The last term on the right is approximated by taking only the term in the summation which has $\mathbf{q}' = -\mathbf{q}$. The terms with other values of $\mathbf{q}'$ are neglected. It is assumed they average out to zero. They do not average to zero, but that is what is meant by the random phase approximation. The approximate equation can now be solved:

$$C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} = \frac{V(\mathbf{q}, t)}{\nu}\left(\frac{C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} - C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}+\mathbf{q},\sigma}}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{q}+\mathbf{q}} + \omega}\right) \tag{5.147}$$

The above equation is now summed over $(\mathbf{p}\sigma)$ to give

$$\rho(\mathbf{q}, t) = \sum_{\mathbf{p}\sigma} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} = \frac{V(\mathbf{q}, t)}{\nu}\sum_{\mathbf{p}\sigma}\frac{C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} - C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}+\mathbf{q},\sigma}}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{q}+\mathbf{q}} + \omega}$$

This equation shows that the operator on the left, $\rho(\mathbf{q}, t)$, is proportional to the operator on the right, $V(\mathbf{q}, t)$. The average is taken of this equation, so that $\langle\rho\rangle$ and $\langle V\rangle$ are replaced by

$\rho(\mathbf{q}, \omega)e^{-i\omega t}$ and $V(\mathbf{q}, \omega)e^{-i\omega t}$. In addition, the number operators $C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}$ and $C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}+\mathbf{q},\sigma}$ are replaced by their averages $n_F(\xi_{\mathbf{p}})$ and $n_F(\xi_{\mathbf{p}+\mathbf{q}})$. These steps give an equation which relates the average of these operators:

$$\rho(\mathbf{q}, \omega) = \frac{V(\mathbf{q}, \omega)}{\nu}\sum_{\mathbf{p}\sigma}\frac{n_F(\xi_{\mathbf{p}}) - n_F(\xi_{\mathbf{p}+\mathbf{q}})}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega} = V(\mathbf{q}, \omega)P^{(1)}(\mathbf{q}, \omega) \tag{5.148}$$

which can now be used in (5.141):

$$V_s(\mathbf{q}, \omega) = \frac{4\pi e^2}{q^2}\rho(\mathbf{q}, \omega) = V(\mathbf{q}, \omega)v_q P^{(1)}(\mathbf{q}, \omega) \tag{5.149}$$

The result has the screening particle density $\rho(\mathbf{q}, \omega)$ proportional to the self-consistent potential $V(\mathbf{q}, \omega)$. The constant of proportionality is $P^{(1)}(\mathbf{q}, \omega)$, which is evaluated below.

The equations may now be solved to obtain the dielectric function. In the equation for $V(\mathbf{q}, \omega)$, substitute the new result for $V_s(\mathbf{q}, \omega)$ and then solve for $V(\mathbf{q}, \omega)$ in terms of $V_i(\mathbf{q}, \omega)$:

$$\begin{aligned} V(\mathbf{q}, \omega) &= V_i(\mathbf{q}, \omega) + V_s(\mathbf{q}, \omega) = V_i(\mathbf{q}, \omega) + v_p P^{(1)}V(\mathbf{q}, \omega) \\ &= \frac{V_i(\mathbf{q}, \omega)}{1 - v_q P^{(1)}(\mathbf{q}, \omega)} \end{aligned} \tag{5.150}$$

The ratio of these two quantities is just the RPA dielectric function:

$$\varepsilon_{\mathrm{RPA}}(\mathbf{q}, \omega) = 1 - v_q P^{(1)}(\mathbf{q}, \omega) \tag{5.151}$$

$$P^{(1)}(\mathbf{q}, \omega) = \frac{1}{\nu}\sum_{\mathbf{p}\sigma}\frac{n_F(\xi_{\mathbf{p}}) - n_F(\xi_{\mathbf{p}+\mathbf{q}})}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega + i\delta} \tag{5.152}$$

The above result completes the derivation using self-consistent fields.

The second method of deriving $\varepsilon_{\mathrm{RPA}}$ is a diagrammatic analysis using Green's functions. The basic definition of $1/\varepsilon(\mathbf{q}, \omega)$ in (5.112) is rewritten in the interaction representation:

$$\frac{1}{\varepsilon(\mathbf{q}, i\omega)} = 1 - \frac{v_q}{\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\frac{\langle T_\tau \hat{S}(\beta)\hat{\rho}(\mathbf{q}, \tau)\hat{\rho}(-\mathbf{q}, 0)\rangle}{\langle \hat{S}(\beta)\rangle} \tag{5.153}$$

$$V = \frac{1}{\nu}\sum_{\mathbf{pkq}}\sum_{\sigma\sigma'} v_q C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C^\dagger_{\mathbf{k}-\mathbf{q},\sigma'}C_{\mathbf{k}\sigma'}C_{\mathbf{p}\sigma} \tag{5.154}$$

where H_0 is the kinetic energy term in the homogeneous electron gas and V is the electron–electron interactions. The operator $\hat{\rho}$ has its time dependence determined by H_0 in the interaction representation. The S matrix will be expanded term by term to see what sort of terms develop. One particular subset of these terms will be summed and will yield the RPA.

The first term in the expansion for the S matrix is just $P^{(1)}(\mathbf{q}, i\omega)$:

$$P^{(1)}(\mathbf{q}, i\omega) = -\frac{1}{\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \hat{\rho}(\mathbf{q}, \tau)\hat{\rho}(-\mathbf{q}, 0)\rangle \tag{5.155}$$

$$= -\frac{1}{\nu}\sum_{\mathbf{kp}}\sum_{\sigma\sigma'}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau)\hat{C}_{\mathbf{k}-\mathbf{q},\sigma'}(0)\hat{C}_{\mathbf{k}\sigma'}(0)\hat{C}_{\mathbf{k}\sigma'}(0)\rangle$$

$$= \frac{1}{\nu}\sum_{\mathbf{p}}\int_0^\beta d\tau e^{i\omega_n\tau}\mathscr{G}^{(0)}(\mathbf{p}, \tau)\mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}, -\tau) \tag{5.156}$$

$$= \frac{1}{\nu}\sum_{\mathbf{p}\sigma}\frac{n_F(\xi_\mathbf{p}) - n_F(\xi_{\mathbf{p}+\mathbf{q}})}{\varepsilon_\mathbf{p} - \varepsilon_{\mathbf{p}+\mathbf{q}} + i\omega} \tag{5.157}$$

The Feynman diagram is shown as a single-fermion closed loop in Fig. 5.9. The wiggly lines at each end are just added to define the two vertices of the polarization diagram. They could indicate that the polarization term is in response to an excitation with wave vector $\mathbf{q}$ and frequency $i\omega_n$. The calculation for ε_{RPA} is not terminated at this point, since the terms so far are $1/\varepsilon = 1 + v_q P^{(1)} + \cdots$ rather than $\varepsilon_{\text{RPA}} = 1 - v_q P^{(1)}$. Obviously, more terms are needed to get RPA.

The next term in the S-matrix expansion is

$$\frac{1}{\nu}\sum_{\mathbf{pkq}'}\sum_{\sigma\sigma'}\int_0^\beta d\tau e^{i\omega_n\tau}\int_0^\beta d\tau_1 v_{q'}$$
$$\times \langle T_\tau \hat{\rho}(\mathbf{q}, \tau)\hat{C}^\dagger_{\mathbf{p}+\mathbf{q}',\sigma}(\tau_1)\hat{C}^\dagger_{\mathbf{k}-\mathbf{q}',\sigma'}(\tau_1)\hat{C}_{\mathbf{k}\sigma'}(\tau_1)\hat{C}_{\mathbf{p}\sigma}(\tau_1)\hat{\rho}(-\mathbf{q}, 0)\rangle \tag{5.158}$$

There are four terms which result when Wick's theorem is applied to this correlation function. All contributions have four electron Green's functions and one Coulomb interaction $v_{q'}$. Their diagrams are shown in Fig. 5.10. The first one is a vertex correction to the basic bubble diagram. The next two are exchange energy diagrams for the Green's functions in the bubble; they contribute to the self-energy of these Green's functions. The last diagram contains two bubbles which are connected by the Coulomb line v_q.

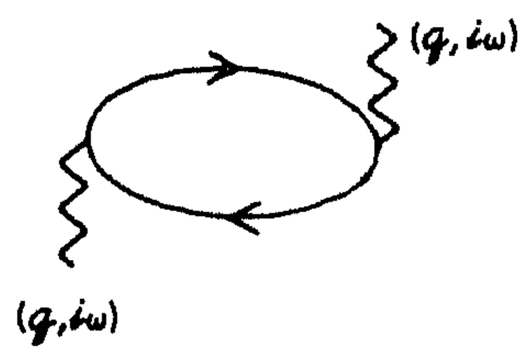


FIGURE 5.9

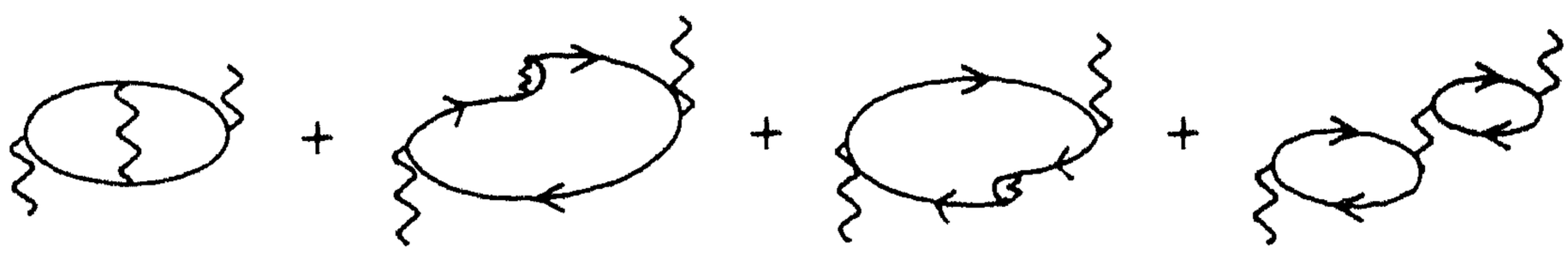


FIGURE 5.10

Earlier it was remarked that the density–density correlation function had the appearance of a Green's function. It also has a Dyson equation. The exact evaluation of this correlation function may be written as

$$-\frac{1}{\nu}\int_0^\beta d\tau e^{i\omega_n T}\langle T_\tau \rho(\mathbf{q},\tau)\rho(-\mathbf{q},0)\rangle = \frac{P(\mathbf{q}, i\omega)}{1 - v_q P(\mathbf{q},\omega)} \tag{5.159}$$

Here the density operators $\rho(\mathbf{q},\tau)$ have the τ dependence governed by $H = H_0 + V$, instead of only H_0 as in $P^{(1)}$. The polarization diagram $P(\mathbf{q}, i\omega)$ is the summation of all "different" polarization terms. Polarization diagrams are not "different" if any of their parts are linked by a single Coulomb line v_q. For example, the last diagram in Fig. 5.10 is not a different polarization diagram. This term arises from the expansion

$$\begin{aligned}\frac{P^{(1)} + P^{(2)} + \cdots}{1 - v_q(P^{(1)} + \cdots)} &= (P^{(1)} + \cdots)[1 + v_q(P^{(1)} + \cdots) + \cdots] \\ &= P^{(1)} + v_q(P^{(1)})^2 + \cdots \end{aligned} \tag{5.160}$$

where it is the term $v_q(P^{(1)})^2$. There are terms in $P(\mathbf{q}, i\omega)$ which have more than one bubble, but they must be connected by more than one Coulomb line.

The random phase approximation is approximating the exact polarization diagram $P(\mathbf{q}, i\omega)$ by its first term, which is $P^{(1)}(\mathbf{q}, i\omega)$. The RPA is found from the summation of all single bubble polarization diagrams

$$\frac{1}{\varepsilon_{RPA}} = 1 + \frac{v_q P^{(1)}}{1 - v_q P^{(1)}} = \frac{1}{1 - v_q P^{(1)}} \tag{5.161}$$

which does give $\varepsilon_{\mathrm{RPA}} = 1 - v_q P^{(1)}(\mathbf{q}, i\omega)$. This derivation makes clear the approximate nature of the RPA. The total polarization operator $P(\mathbf{q}, i\omega)$ has an infinite number of terms, while the RPA retains one. The exact dielectric function is easily shown to be

$$\varepsilon(\mathbf{q}, i\omega) = 1 - v_q P(\mathbf{q}, i\omega) \tag{5.162}$$

An obvious way to improve the dielectric function is to include more terms in the summation of polarization contributions (Geldart and Taylor, 1970). This step is not as simple as it sounds. There are an infinite number of possibilities, so some physics must be used to guide the choice. Neither does it help that the obvious possibilities, such as the first diagram in Fig. 5.10, are not simple to evaluate analytically. In fact, most progress has been made by nondiagrammatic means, as will be discussed below.

The RPA dielectric function is evaluated once $P^{(1)}(\mathbf{q}, i\omega)$ is obtained. The retarded function is obtained by taking the analytical continuation $i\omega \to \omega + i\delta$. The retarded dielectric function is complex, and its real and imaginary parts are called ε_1 and ε_2:

$$\varepsilon_{\mathrm{RPA}}(\mathbf{q},\omega) = \varepsilon_1(\mathbf{q},\omega) + i\varepsilon_2(\mathbf{q},\omega) \tag{5.163}$$

$$\begin{aligned}\varepsilon_1(\mathbf{q},\omega) = 1 + \frac{q_{TF}^2}{2q^2}\Bigg\{ &1 + \frac{m^2}{2k_F q^3}[4E_F\varepsilon_q - (\varepsilon_q + \omega)^2]\ln\left|\frac{\varepsilon_q + qv_F + \omega}{\varepsilon_q - qv_F + \omega}\right| \\ &+ \frac{m^2}{2k_F q^3}[4E_F\varepsilon_q - (\varepsilon_q - \omega)^2]\ln\left|\frac{\varepsilon_q + qv_F - \omega}{\varepsilon_q - qv_F - \omega}\right|\Bigg\}\end{aligned} \tag{5.164}$$

• For $q < 2k_F$

$$\begin{aligned}
\varepsilon_2(\mathbf{q}, \omega) &= 2\omega\left(\frac{e^2m^2}{q^3}\right), && qv_F - \varepsilon_q > \omega > 0 \\
&= \left(\frac{e^2m^2}{q^3}\right)\left[k_F^2 - \left(\frac{m}{q}\right)^2(\omega - \varepsilon_q)^2\right], && \varepsilon_q + qv_F > \omega > qv_F - \varepsilon_q \\
&= 0 && \omega > \varepsilon_q + qv_F
\end{aligned} \tag{5.165}$$

• For $q > 2k_F$

$$\varepsilon_2(\mathbf{q}, \omega) = \left(\frac{e^2m^2}{q^3}\right)\left[k_F^2 - \left(\frac{m}{q}\right)^2(\omega - \varepsilon_q)^2\right], \qquad \varepsilon_q + qv_F > \omega > qv_F - \varepsilon_q$$

The real part $\varepsilon_1(\mathbf{q}, \omega)$ may be represented by a single formula. The imaginary part $\varepsilon_2(\mathbf{q}, \omega)$ has a variety of functional relations for different values of $(\mathbf{q}, \omega)$.

Figure 5.11 shows graphs for $\varepsilon_1, \varepsilon_2$, and $-\mathrm{Im}(1/\varepsilon)$ for $r_s = 3$ and all relevant frequencies. For large values of q, say $q \geq 2k_F$, one finds that $\varepsilon_1 \approx 1$, ε_2 is small, so that $\varepsilon_2 \approx -\mathrm{Im}(1/\varepsilon)$. There is only a single line drawn for these cases.

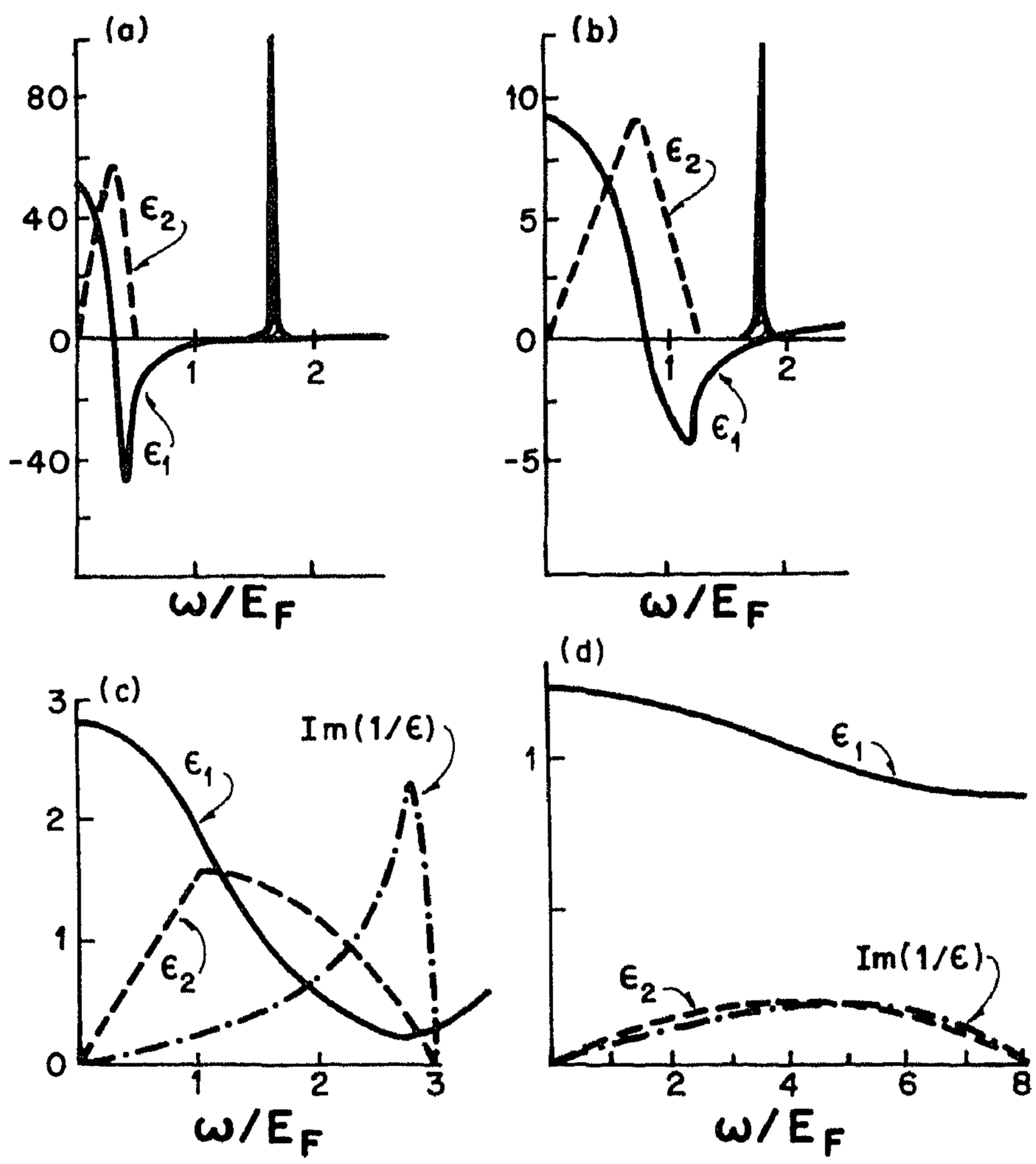


FIGURE 5.11 The RPA predictions regarding ε_1, ε_2, $\mathrm{Im}(1/\varepsilon)$ as a function of q and ω. Results shown for $r_s = 3.0$.

The real part ε_1 always approaches 1 at large ω. At $\omega \to 0$ then $\varepsilon_2 \to 0$, and the static ε_{RPA} is just $\varepsilon_1(\mathbf{q}, 0)$. Using the notation ($x = q/2k_F$),

$$\varepsilon(\mathbf{q}, 0) = 1 + \frac{q_{TF}^2}{2q^2}\left[1 + \frac{1}{2x}(1 - x^2)\ln\left|\frac{1+x}{1-x}\right|\right] \tag{5.166}$$

which is always positive. The prefactor to the bracket includes the square of the Thomas–Fermi wave vector q_{TF}. For values of $q < k_F$, then $\varepsilon_1(q, \omega)$ becomes negative for intermediate values of ω. This requires two crossings of the $\varepsilon_1 = 0$ axis. The low-frequency crossing always happens when ε_2 is large, so that $-\text{Im}(1/\varepsilon) = \varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2)$ is well behaved when $\varepsilon_1 \approx 0$. However, the high-frequency point where $\varepsilon_1 = 0$ has $\varepsilon_2 = 0$. In that case $-\lim(1/\varepsilon) = \varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2) = \pi\delta(\varepsilon_1)$, so that a delta function is obtained. This delta function is the plasmon peak which is the sharp singularity on the right of the graph. It is given a nonzero width in Fig. 5.11 to aid the eye. Remember that peaks in $S(q, \omega)$ are interpreted as excitations of the system

$$S(\mathbf{q}, \omega) = -\frac{1}{n_0 v_q}\text{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right] = \frac{1}{n_0 v_q}\frac{\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2} \tag{5.167}$$

Plasmons are excitations which exist in real metals and in any electron gas. They were discovered in ionized gases by Langmuir. They occur at small values of q, as is evident in Fig. 5.11.

Examine the limit of (5.164) when $q \to 0$ while ω remains nonzero. The first step is to expand $P^{(1)}(\mathbf{q}, \omega)$ while assuming $\varepsilon_q \ll \omega$, $qv_F \ll \omega$. Let $\Lambda = \varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}$ and assume that $\Lambda < \omega$ so that

$$P^{(1)}(\mathbf{q}, \omega) = \frac{2}{\nu}\sum_{\mathbf{p}} n_F(\xi_{\mathbf{p}})\left[\frac{1}{\omega + \varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}}} - \frac{1}{\omega - \varepsilon_{\mathbf{p}} + \varepsilon_{\mathbf{p}+\mathbf{q}}}\right] \tag{5.168}$$

$$= \frac{2}{\nu\omega}\sum_{\mathbf{p}} n_F(\xi_{\mathbf{p}})\left[\frac{1}{1 - \dfrac{\Lambda}{\omega}} - \frac{1}{1 + \dfrac{\Lambda}{\omega}}\right] \approx \frac{4}{\nu\omega}\sum_{\mathbf{p}} n_F(\xi_{\mathbf{p}})\left[\frac{\Lambda}{\omega} + \frac{\Lambda^3}{\omega^3} + \cdots\right]$$

$$= \frac{2n_0\varepsilon_{\mathbf{q}}}{\omega^2}\left[1 + \frac{1}{\omega^2}\left[\tfrac{3}{5}(qv_F)^2 + \varepsilon_{\mathbf{q}}^2\right] + \cdots\right] \tag{5.169}$$

This result is inserted into $\varepsilon = 1 - v_q P^{(1)}$ which derives the high frequency limit of the dielectric function

$$\lim_{q\to 0}\varepsilon_1(\mathbf{q}, \omega) = 1 - \frac{\omega_p^2}{\omega^2}\left\{1 + \frac{1}{\omega^2}\left[\tfrac{3}{5}(qv_F)^2 + \varepsilon_q^2\right] + O\left(\frac{1}{\omega^4}\right)\right\} \tag{5.170}$$

$$\omega_0^2 = \frac{4\pi e^2 n_0}{m} \tag{5.171}$$

For $\varepsilon_2 = 0$ the condition that $\varepsilon_1 = 0$ predicts that the plasmon peak occurs at the frequency

$$\omega = \omega_p\left(1 + \frac{3}{10}\frac{q^2 v_F^2}{\omega_p^2} + \cdots\right) \tag{5.172}$$

At $q = 0$ the prediction is quite simple, $\omega = \omega_p$. The quantity ω_p is called the *plasma frequency* of the electron gas. It depends only on the electron density n_0 and mass m, which is actually the effective band mass, although these are close to the free-electron value for many metals. Table 5.1 shows some actual plasma frequencies measured in metals, which are

TABLE 5.1 Plasma frequencies in eV. The values for $\hbar\omega_p'$ are corrected for the dielectric screening of the core electrons

Metal	Experimental	$\hbar\omega_p$	$\hbar\omega_p'$	Reference
Li	7.12	8.03	7.95	*a*
Na	5.71	5.90	5.77	*a*
	5.85			*b*
K	3.72	4.36	4.02	*a*
	3.87			*c*
Mg	10.6	10.88	10.70	*d*
Al	15.3	15.77	15.55	*e*

[a] C. Kunz, *Phys. Lett.* **15**, 312 (1965).
[b] J. B. Swan, *Phys. Rev.* **135**, A1467 (1964).
[c] J. L. Robins and F. E. Best, *Proc. Phys. Soc. London* **79**, 110 (1962).
[d] C. J. Powell and J. B. Swan, *Phys. Rev.* **116**, 81 (1959).
[e] C. J. Powell and J. B. Swan, *Phys. Rev.* **115**, 869 (1959).

compared with this simple formula. The predictions for ω_p are found to be amazingly accurate. One great reason for the popularity of the RPA dielectric function is that it obviously describes plasmons very well. Later it is shown that it is exact in the limit where $q \to 0$, so that this success is understandable.

The formulas for $\varepsilon_2(\mathbf{q}, \omega)$ in (5.165) are complicated. They may be derived in the following fashion. First, begin with the definition of the imaginary part of the retarded function:

$$\begin{aligned}\varepsilon_2(\mathbf{q}, \omega) &= -v_q \operatorname{Im}[P^{(1)}_{\text{ret}}(\mathbf{q}, \omega)] \\ &= 2\pi v_q \int \frac{d^3p}{(2\pi)^3} \delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega)[n_F(\xi_{\mathbf{p}}) - n_F(\xi_{\mathbf{p}+\mathbf{q}})]\end{aligned} \tag{5.173}$$

A variable change is made in the term with $n_F(\xi_{\mathbf{p}+\mathbf{q}})$ by replacing $\mathbf{p} \to -\mathbf{p} - \mathbf{q}$, so the above equation becomes

$$\begin{aligned}\varepsilon_2(\mathbf{q}, \omega) &= \frac{v_q}{(2\pi)^2} \int d^3p n_F(\xi_{\mathbf{p}})[\delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega) - \delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} - \omega)] \\ &= \frac{v_q}{2\pi} \int_0^{k_F} p^2 dp \int_{-1}^{1} d\nu \left[\delta\left(\varepsilon_q + \frac{qp\nu}{m} - \omega\right) - \delta\left(\varepsilon_q + \frac{qp\nu}{m} + \omega\right)\right]\end{aligned} \tag{5.174}$$

This formula shows that $\varepsilon_2(\mathbf{q}, \omega)$ is antisymmetric in frequency $\varepsilon_2(\mathbf{q}, \omega) = -\varepsilon_2(\mathbf{q}, -\omega)$. It will be evaluated first for $\omega > 0$. The angular integrations ($\nu = \cos\theta$) are done first, and they eliminate the delta function:

$$\int_{-1}^{1} d\nu \delta\left(\varepsilon_q + \frac{qp\nu}{m} \pm \omega\right) = \frac{m}{pq} \theta(p - p_{1,2}) \tag{5.175}$$

$$p_{1,2} = \frac{m}{q} |\omega \pm \varepsilon_q| \tag{5.176}$$

$$\varepsilon_2(q, \omega) = \frac{2me^2}{q^3} \left(\int_{p_1}^{k_F} p dp - \int_{p_2}^{k_F} p dp \right) \tag{5.177}$$

$$\varepsilon_2(q, \omega) = \frac{me^2}{q^3} [\theta(k_F - p_1)(k_F^2 - p_1^2) - \theta(k_F - p_2)(k_F^2 - p_2^2)] \tag{5.178}$$

The complicated aspect of the integral comes from the lower limits of integration p_1 and p_2. These limits are imposed by the angular integral over the delta function. For example, in the first integral the delta function forces $\nu = \cos\theta$ to equal

$$-1 \leq \nu = (\omega - \varepsilon_q)\frac{m}{pq} \leq 1 \tag{5.179}$$

and ν must have values between ± 1. The latter conditions restrict the value of p to

$$-p \leq \frac{m}{q}(\omega - \varepsilon_q) \leq p \tag{5.180}$$

$$p \geq \frac{m}{q}|\omega - \varepsilon_q| = p_1 \tag{5.181}$$

The integral contributes only when $k_F^2 \geq p^2 \geq p_1^2$. In the second integral, the result is similar and only the sign of ω is changed. The result in (5.178) is identical with (5.165), although there the result is simplified. For example, $k_F > p_2$ may be satisfied only when $q < 2k_F$, so the second term may be eliminated when $q > 2k_F$.

At small values of q, $\varepsilon_2(q, \omega)$ is proportional to ω at small values of ω. This dependence is shown in (5.165) and is evident from the graphs in Fig. 5.11. The proportionality to ω is an important feature of ε_2. It arises in (5.178) whenever both inequalities $k_F > p_1, p_2$ are satisfied, since $p_1^2 - p_2^2 = 2m\omega$. The linear dependence of ε_2 on ω must also occur for the exact dielectric function, which may be shown by a simple argument. The physical process under consideration is the rate at which electron–hole pairs are made in the electron gas. A *hole* is a state which has an electron removed from the filled Fermi sea. An initial electron of momentum $\mathbf{p}$ and energy $\xi_{\mathbf{p}}$ is excited by a perturbation with $(\mathbf{q}, \omega)$. The electron is excited to a new state with momentum $\mathbf{p} + \mathbf{q}$ and energy $\xi_{\mathbf{p}+\mathbf{q}} = \xi_{\mathbf{q}} + \omega$. The electron can only be scattered into states which are previously unoccupied, so $\xi_{\mathbf{p}+\mathbf{q}}$ must be above the occupied Fermi sea. The basic process takes an electron from below to above the Fermi level. It leaves a vacancy in the Fermi sea, which is the hole. The excitation process makes electron–hole pairs. The net rate of making such pairs is

$$R_{\text{pairs}} = 4\pi \int \frac{d^3p}{(2\pi)^3}\delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega)\{n_F(\xi_{\mathbf{p}})[1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})] - n_F(\xi_{\mathbf{p}+\mathbf{q}})[1 - n_F(\xi_{\mathbf{p}})]\} \tag{5.182}$$

where $n_F(\xi_{\mathbf{p}})[1 - n_F(\xi_{\mathbf{p}+\mathbf{q}})]$ is the rate of making pairs, while $n_F(\xi_{\mathbf{p}+\mathbf{q}})[1 - n_F(\xi_{\mathbf{p}})]$ is the return rate. A formula similar to (5.173) is derived, except that neither $\xi_{\mathbf{p}}$ nor $n_F(\xi_{\mathbf{p}})$ need to be their free-particle values. They could be interpreted as the energy and occupation number for the fully interacting system. Now change integration variables to $pdp = md\xi_p \equiv md\xi$: and $pd\nu = (m/q)d\xi_{\mathbf{p}+\mathbf{q}} = (m/q)d\xi'$, the above integral may be written as

$$R_{\text{pairs}} \approx \frac{m^2}{\pi q}\int d\xi \int d\xi'\delta(\xi + \omega - \xi')[n_F(\xi) - n_F(\xi')] \tag{5.183}$$

$$\approx \frac{m^2}{\pi q}\int d\xi[n_F(\xi) - n_F(\xi + \omega)] \tag{5.184}$$

where the limits of integration are ignored in this simple argument. The remaining integral is over a distribution $n_F(\xi)$ minus the same integral over the displaced distribution $n_F(\xi + \omega)$. The difference between these two must be proportional to ω, at small ω, since

$$n_F(\xi + \omega) \approx n_F(\xi) + \omega \frac{dn_F}{d\xi} \tag{5.185}$$

$$R_{\text{pairs}} \approx \frac{\omega m^2}{\pi q} \int d\xi \left[-\frac{dn_F(\xi)}{d\xi} \right] = \frac{\omega m^2}{\pi q} \tag{5.186}$$

The rate of making electron–hole pairs is proportional to ω. The derivation applies only to small values of ω, since at large values the limits of integration change the result as they did in the case of the RPA. The area in (q, ω) space over which $\varepsilon_2(q, \omega)$ is nonzero is shown in Fig. 5.12 by the shaded region. These are bounded by the two lines $\omega = \varepsilon_q \pm q v_F$. The excitation spectrum of the electron gas is given by $S(q, \omega)$ in (5.167). It is nonzero when ε_2 is nonzero, so that it also exists in the shaded area of Fig. 5.12. In addition, $S(q, \omega)$ has the plasmon peak which exists where $\varepsilon_2 = 0$. It is also shown in Fig. 5.12. This figure actually describes the excitation spectrum of the electron gas to density fluctuations. At large values of q/k_F, $S(q, \omega)$ is given accurately by the approximation $S(q, \omega) \sim \varepsilon_2/n_0 v_q$, so that it describes the process whereby single electron–hole pairs are made. At intermediate values of q/k_F, the excitation spectrum has the complex shape shown in Fig. 5.12. These curves can be measured in metals by the inelastic scattering of energetic electrons through large momentum transfers. However,the comparison with experiments is deferred until later sections, after several other dielectric models have been introduced.

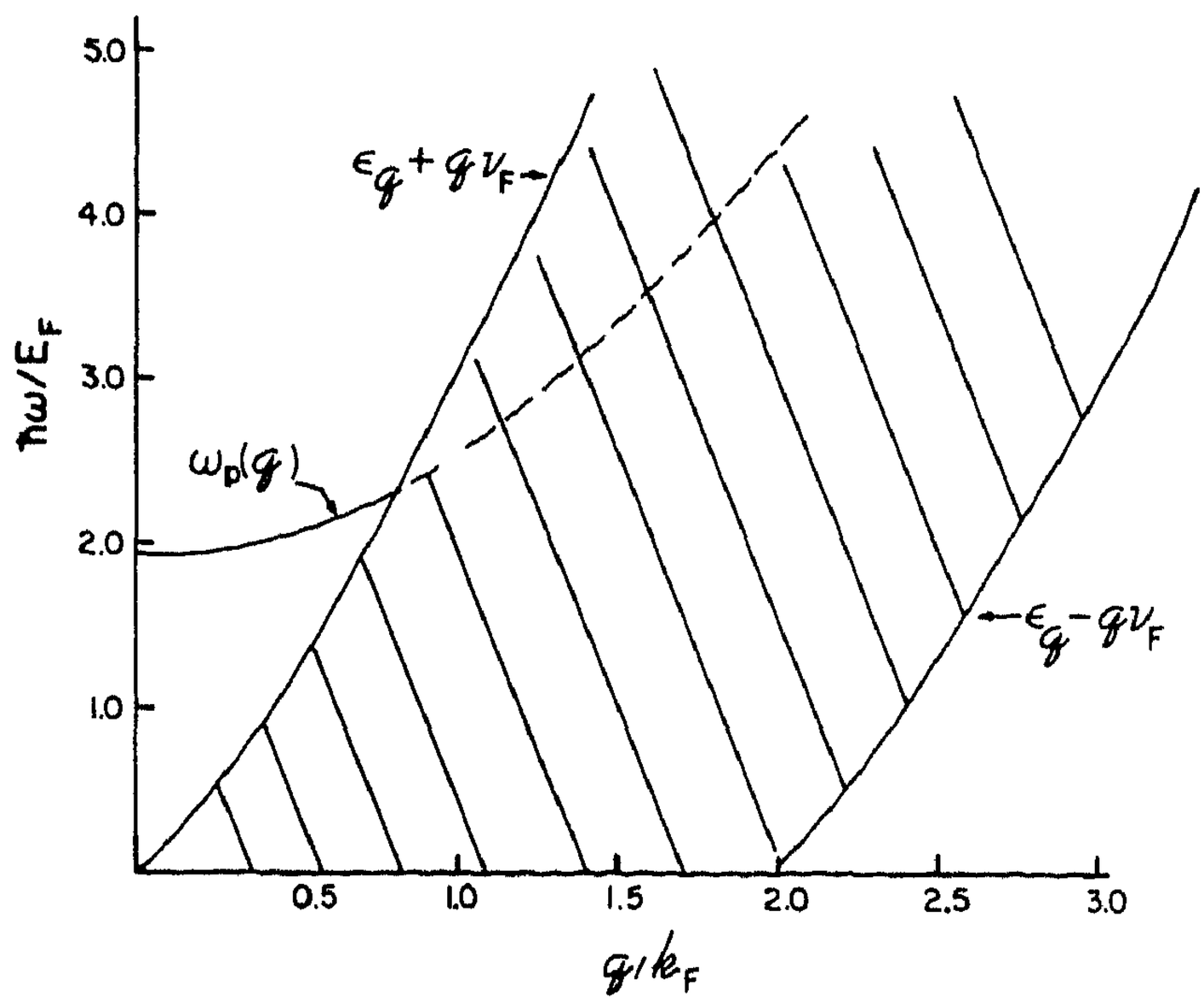


FIGURE 5.12 The excitation region of the electron gas in $(\mathbf{q}, \omega)$ space. The plasmon line ω_p becomes highly damped in the region of electron–hole pairs, which is shown hatched.

5.5.3. Hubbard

Hubbard (1957) introduced a correction factor to the RPA of the form

$$\varepsilon_H(\mathbf{q}, \omega) = 1 - \frac{v_q P^{(1)}(\mathbf{q}, \omega)}{1 + v_q G_H(q) P^{(1)}(\mathbf{q}, \omega)} \tag{5.187}$$

$$G_H(q) = \frac{1}{2}\frac{q^2}{q^2 + k_F^2} \tag{5.188}$$

The factor $G_H(q)$ is a local field correction. There is an analogy with the Lorenz–Lorentz dielectric function, which has the form in cubic insulators

$$\varepsilon = 1 + \frac{4\pi\alpha}{1 - \dfrac{4\pi\alpha}{3}} \tag{5.189}$$

The two formulas have the same structure if we associate $-v_q P^{(1)} \Rightarrow 4\pi\alpha$ and $G_H \Rightarrow \frac{1}{3}$. The factor of $4\pi\alpha/3$ comes from summing the local dipoles in the lattice. It is called a "local field correction". Similarly, $G_H(q)$ in the Hubbard dielectric function comes from summing over the surrounding structure of the dielectric, which in this case is the exchange-correlation hole.

This curious formula was regarded as an improvement to the RPA in many properties. Yet its true worth was unappreciated, because it could not be compared with the better dielectric functions which only became available later. The best dielectric functions today are written in precisely the form (5.187), with a $G_H(q)$ which is only slightly different from the simple form proposed by Hubbard. His result, which was somewhat of a stab in the dark, is now regarded as being well ahead of its time.

The factor $G_H(q)$ is introduced to account for the existence of the exchange and correlation hole around the electron. The dielectric function describes how the conduction electrons of the metal rearrange their positions to screen the Coulomb potential. Because of the exchange and correlation hole around each electron, when one electron is participating in the dielectric screening, others are less likely to be found nearby. The exchange-correlation hole should have some affect upon the nature of the dielectric screening.

The Hubbard local field comes from the vertex corrections in the polarization function $P(\mathbf{q}, \omega)$. The most important vertex corrections are the ladder diagrams in Fig 5.13. They can be evaluated by introducing a vertex function $\Gamma(k, k+q)$ in a four vector notation $k \equiv (\mathbf{k}, ik_n)$. This approximation to the polarization diagrams is denoted P_L where "L" denotes ladder

$$\begin{aligned} P_L(\mathbf{q}, \omega) &= \frac{2}{\beta\nu}\sum_k \mathcal{G}^{(0)}(k)\mathcal{G}^{(0)}(k+q)\Gamma(k, k+q) \\ \Gamma(k, k+q) &= 1 - \frac{1}{\beta\nu}\sum_p W(k-p)\mathcal{G}^{(0)}(p)\mathcal{G}^{(0)}(p+q)\Gamma(p, p+q) \end{aligned} \tag{5.190}$$

The above expression is rather difficult to evaluate exactly. The degree of complexity depends upon the choice for the effective interaction $W(k-p)$ which are the ladders.

In the Hartree–Fock approximation, to the ladder diagrams, the function $W(k-p)$ is the unscreened Coulomb interaction $W(q) = v_q$. Since the interaction has no dependence upon

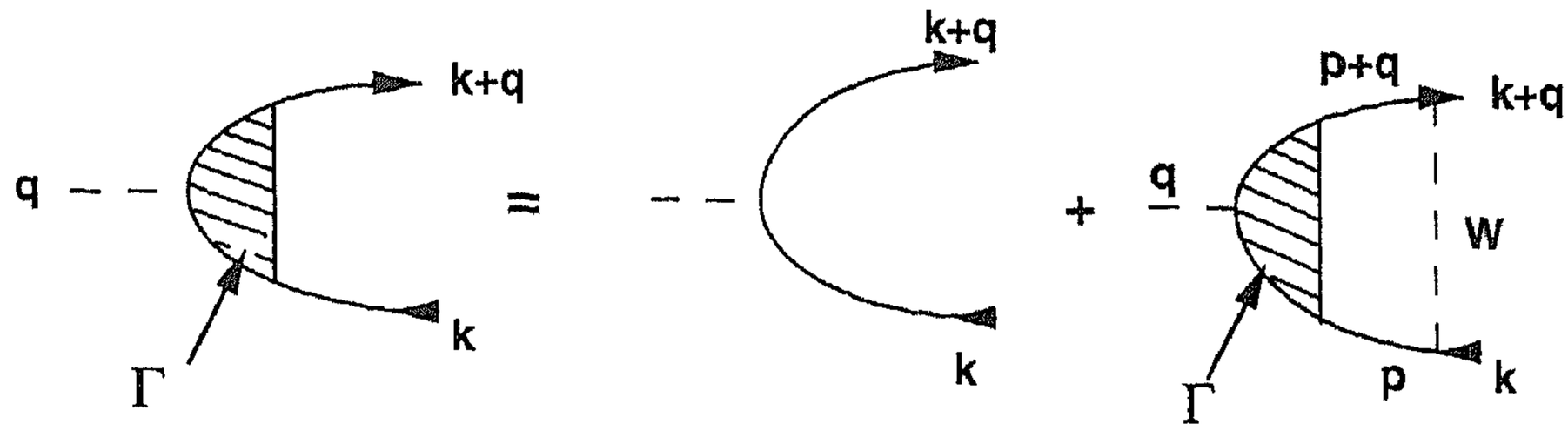


FIGURE 5.13 Ladder diagrams of the polarization bubble.

frequency, neither does the vertex function $\Gamma(\mathbf{k}, \mathbf{k}+\mathbf{q})$ depend upon ik_n. In this special case the frequency summation ip_n is simple and the above expression is

$$\Gamma(\mathbf{k}, \mathbf{k}+\mathbf{q}: iq_m) = 1 - \int \frac{d^3p}{(2\pi)^3} v_{|\mathbf{k}-\mathbf{p}|} \Gamma(\mathbf{p}, \mathbf{p}+\mathbf{q}: iq_m) \frac{n_{\mathbf{p}} - n_{\mathbf{p}+\mathbf{q}}}{iq_m + \varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}}}$$

This expression has been evaluated by Sham (1973) and by Brosens *et al.* (1977). They also added the exchange self-energies $\Sigma_x(\mathbf{p}) - \Sigma_x(\mathbf{p}+\mathbf{q})$ to the energy denominator. In this case the theory is called "conserving" (Baym and Kadanoff, 1961) in that the vertex corrections and self-energies are both evaluated to the same level of approximation. This dielectric function has some interesting structure.

A better approximation is to have $W(q)$ be a screened interaction such as $W(q) = v_q/\varepsilon(q)$. In this case the dielectric function $\varepsilon(q)$ depends upon frequency (iq_m) which complicates the evaluation of the integral equation for the ladder diagrams. An approximate result is obtained by making some drastic approximations. The first one is to assume that $\Gamma(k, k+q) \approx \Gamma(q)$ and the k dependence is ignored. This approximation signficantly reduces the complexity of the analysis, since the equations now have the form

$$P_L(q) = \Gamma(q)P^{(1)}(q) \tag{5.191}$$

$$\Gamma(q) = 1 - \Gamma(q)\Lambda(q) = \frac{1}{1+\Lambda(q)} \tag{5.192}$$

$$\Lambda(q) = \frac{1}{\beta\nu}\sum_p W(k-p)\mathscr{G}^{(0)}(p)\mathscr{G}^{(0)}(p+q) \tag{5.193}$$

There is a problem with the last expression. If $\Gamma(q)$ does not depend upon (q), then neither can $\Lambda(q)$. Therefore, the factor of $W(k-p)$ must have its argument replaced by another one. Here is where Hubbard made his important approximation. He replaced $W(k-p)$ by a statically screened interaction

$$W(k-p) \Rightarrow \frac{4\pi e^2}{q^2 + q_{TF}^2} \tag{5.194}$$

$$\Lambda(q) = \frac{2\pi e^2}{q^2 + q_{TF}^2} P^{(1)}(q) = v_q G_H(q) P^{(1)}(q) \tag{5.195}$$

$$G_H(q) = \frac{1}{2}\frac{q^2}{q^2 + q_{TF}^2} \tag{5.196}$$

Once the factor of $W(q)$ is removed from the summation over p, then that summation just gives $\frac{1}{2}P^{(1)}(q)$. The factor of $\frac{1}{2}$ comes because $P^{(1)}$ contains a summation over spins which

multiplies the result by two. The ladder terms do not have a summation over spin states. Each electron line in the polarization bubble diagram has the same spin configuration. So the result is multiplied by $\frac{1}{2}$ to account for the factor of 2 in $P^{(1)}$. Note that the derivation gave a Hubbard factor with a denominator of $(q^2 + q_{TF}^2)$ rather than the original Hubbard expression of $(q^2 + k_F^2)$. Actually, Hubbard waffled, and published both results in separate papers. Another time he suggested

$$G_H(q) = \frac{1}{2}\frac{q^2}{q^2 + k_F^2 + q_{TF}^2} \tag{5.197}$$

This latter result is remarkably close to the best theories today.

Why do the ladder diagrams come from the exchange-correlation hole? The polarization diagram is a bubble with two electron lines. One is going forward in time ("electron") and one is going backward in time ("hole"). The hole is an excitation from below the chemical potential. The ladder diagrams introduce scattering, and therefore correlations, between the electrons and holes. The correlation hole around the electron must be created by the low energy excitations of the electron gas, which are just the electron–hole pairs. The correlation hole is due to correlations between electrons and holes.

5.5.4. Singwi–Sjölander

The 1967 Singwi–Sjölander dielectric function has the same form (5.187) as Hubbard's but with a different choice of $G_H(q)$. Singwi has collaborated with a variety of authors to develop improvements in the method of choosing $G_H(q)$. Once could remark, in a humorous way, that $G_H(q)$ is time dependent because of its improvements over the years. This variety makes the discussion difficult, since there are several possible choices to describe. The original version of Singwi–Sjölander is derived, and the others are cited in the references (see Vashishta and Singwi, 1972; Singwi and Tosi, 1981).

Their original derivation is very attractive, because it explicitly includes the exchange-correlation hole. They derive an equation of motion for the screening charge and then insert the pair distribution function $g(r)$ into the Coulomb integral between particles. They solved these equations and obtained

$$\begin{aligned} \varepsilon(\mathbf{q}, \omega) &= 1 - \frac{v_q P^{(1)}(\mathbf{q}, \omega)}{1 + v_q G_H(q) P^{(1)}(\mathbf{q}, \omega)} \\ G_H(q) &= -\frac{1}{n_0}\int \frac{d^3k}{(2\pi)^3}\frac{\mathbf{q}\cdot\mathbf{k}}{k^2}[S(\mathbf{q}-\mathbf{k}) - 1] \end{aligned} \tag{5.198}$$

where $S(q)$ is the static structure factor associated with $g(r)$. Their original derivation employs Wigner distribution functions. An alternate derivation is presented here which was given by Singwi *et al.* (1968). It uses equations of motion. They find the second time derivative of the particle density operator $\rho(\mathbf{q}, t)$ and derive an expression for the plasma frequency. The plasma frequency is given by $\varepsilon(q, \omega) = 0$. Remember that at high frequency $v_q P^{(1)} \approx \omega_p^2/\omega^2$, then using the equation for the dielectric function gives

$$\varepsilon = 0 = 1 - \frac{\omega_p^2}{\omega_2}\left[1 - G_H(q) + O\left(\frac{q^2 v_F^2}{\omega^2}\right)\right] \tag{5.199}$$

Corrections to the plasma frequency can be associated with the Hubbard local field G_H.

The first time derivative is a single commutator with H, so the second time derivative is a double commutator,

$$\ddot{\rho}(\mathbf{q}, t) = -[H, [H, \rho(\mathbf{q}, t)]] \tag{5.200}$$

$$\rho(\mathbf{q}, t) = \sum_{\mathbf{p}\sigma} C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{5.201}$$

where H is the Hamiltonian for the homogeneous electron gas. Consider the first commutator,

$$[H, \rho(\mathbf{q}, t)] = \sum_{\mathbf{p}\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}) C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{5.202}$$

where $\rho(\mathbf{q}, t)$ commutes with the electron–electron term in the Hamiltonian. That is not surprising, since electron–electron interactions can be written as $\sum_{\mathbf{q}} v_q \rho(\mathbf{q})\rho(-\mathbf{q})$ except for self-interaction effects. An operator commutes with itself. The electron–electron term contains just density operators, which is why it commutes with $\rho(\mathbf{q}, t)$. The second commutator provides the interesting result. Another commutator with the kinetic term in H just gives a repetition of the energy difference factor:

$$\left[\sum_{\mathbf{k},s} \xi_{\mathbf{k},s} C^{\dagger}_{\mathbf{k},s} C_{\mathbf{k},s}, \sum_{\mathbf{p}\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}) C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}\right] = \sum_{\mathbf{p}\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}})^2 C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}$$

The commutator with the interaction gives four terms, which can be combined to give two different ones by rearranging the order of operators. They are

$$\begin{aligned}
&\frac{1}{2\nu} \sum_{\mathbf{k}\mathbf{p}\mathbf{p}'\mathbf{q}'} \sum_{s,s',\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}) v_{q'} [C^{\dagger}_{\mathbf{p}'+\mathbf{q}',s} C^{\dagger}_{\mathbf{k}+\mathbf{q}',s'} C_{\mathbf{k},s'} C_{\mathbf{p}'s}, C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}] \\
&\quad = \frac{1}{\nu} \sum_{k\mathbf{p}\mathbf{q}'s\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}) v_{q'} [C^{\dagger}_{\mathbf{p}+\mathbf{q}+\mathbf{q}'\sigma} C_{\mathbf{k}-\mathbf{q}',s} C_{\mathbf{k}s} C_{\mathbf{p}\sigma} - C^{\dagger}_{\mathbf{p}+\mathbf{q}\sigma} C^{\dagger}_{\mathbf{k}-\mathbf{q}',s} C_{\mathbf{k}s} C_{\mathbf{p}-\mathbf{q}',\sigma}]
\end{aligned} \tag{5.203}$$

The change in dummy variables $\mathbf{p} \to \mathbf{p} + \mathbf{q}'$ in the last term makes the operator sequence the same as the first term:

$$= \frac{1}{\nu} \sum_{k\mathbf{p}\mathbf{q}'s\sigma} v_{q'} C^{\dagger}_{\mathbf{p}+\mathbf{q}+\mathbf{q}'\sigma} C^{\dagger}_{\mathbf{k}-\mathbf{q}',s} C_{\mathbf{k}s} C_{\mathbf{p}\sigma} [(\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}) - (\varepsilon_{\mathbf{p}+\mathbf{q}+\mathbf{q}'} - \varepsilon_{\mathbf{p}+\mathbf{q}'})]$$

The kinetic energies largely cancel in this expression. The bracket is simply $-\mathbf{q}\cdot\mathbf{q}'/m$. In writing this term, the operator sequence $C^{\dagger}_{\mathbf{p}+\mathbf{q}+\mathbf{q}'\sigma} C^{\dagger}_{\mathbf{k}-\mathbf{q}',s} C_{\mathbf{k}s} C_{\mathbf{p}\sigma}$ will be rearranged into $C^{\dagger}_{\mathbf{p}+\mathbf{q}+\mathbf{q}'\sigma} C_{\mathbf{p}\sigma} C^{\dagger}_{\mathbf{k}-\mathbf{q}',s} C_{\mathbf{k}s}$. After summing over internal variables, this expression becomes $\rho(\mathbf{q}+\mathbf{q}', t)\rho(-\mathbf{q}', t)$. The rearrangement is permissible as long as the term is omitted which has a particle acting on itself; a self-interaction is included in $\rho\rho$ unless care is taken to leave it out. The various terms are collected into the equation of motion for the density operator:

$$-\ddot{\rho}(\mathbf{q}, t) = \sum_{\mathbf{p}\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}})^2 C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{5.204}$$

$$+ \frac{1}{\nu} \sum_{\mathbf{q}'} v_{q'} \left(\frac{-\mathbf{q}\cdot\mathbf{q}'}{m}\right) \rho(\mathbf{q}+\mathbf{q}', t)\rho(-\mathbf{q}', t) \tag{5.205}$$

An obvious way to evaluate the last term is to take $\mathbf{q}' = -\mathbf{q}$ in the summation. This term alone gives the equation $[\rho(0) = N_e]$

$$-\ddot{\rho}(\mathbf{q}, t) = \frac{v_q q^2}{m} \frac{N_e}{\nu} \rho(\mathbf{q}, t) = \omega_p^2 \rho(\mathbf{q}, t) \tag{5.206}$$

which describes the plasma oscillations in the electron gas. Since the other terms all vanish when $\mathbf{q} \to 0$, it has been shown that the long-wavelength density oscillations are the same plasma oscillations as predicted in the RPA. The RPA is exact in this limit.

The Singwi–Sjölander result is obtained from treating $\langle \rho(\mathbf{q}+\mathbf{q}', t)\rho(-\mathbf{q}', t)\rangle$ more carefully. They wrote it in terms of a direct summation over particle positions, as was done in Sec. 1.6:

$$\rho(\mathbf{q}+\mathbf{q}')\rho(-\mathbf{q}') = \sum_i e^{\mathbf{r}_i \cdot (\mathbf{q}+\mathbf{q}')} \sum_{j \neq i} e^{-i\mathbf{r}_j \cdot \mathbf{q}'} \tag{5.207}$$

The term $i = j$ is excluded, since it is a particle interacting with itself. The factors in the exponent are rearranged to have one summation over a particle $\mathbf{r}_j$ and another summation over the particle separations $\mathbf{r}_i - \mathbf{r}_j$:

$$\rho(\mathbf{q}+\mathbf{q}')\rho(-\mathbf{q}') = \sum_j e^{i\mathbf{r}_j \cdot \mathbf{q}} \sum_{i \neq j} e^{i(\mathbf{r}_i - \mathbf{r}_j)\cdot(\mathbf{q}+\mathbf{q}')} \tag{5.208}$$

By following the definitions in Sec. 1.6, the summation over $\mathbf{r}_i - \mathbf{r}_j$ can be replaced by its average in the electron gas:

$$\sum_{i \neq j} e^{i(\mathbf{r}_i - \mathbf{r}_j)\cdot(\mathbf{q}+\mathbf{q}')} \Rightarrow n_0 \int d^3r g(r) e^{i\mathbf{r}\cdot(\mathbf{q}+\mathbf{q}')} = N_e \delta_{\mathbf{q}+\mathbf{q}'} + S(\mathbf{q}+\mathbf{q}') - 1$$

The other summation $\sum e^{i\mathbf{q}\cdot\mathbf{r}_j}$ is just $\rho(\mathbf{q}, t)$, so that

$$\langle \rho(\mathbf{q}+\mathbf{q}')\rho(-\mathbf{q}')\rangle' = \langle \rho(\mathbf{q}, t)\rangle [N_e \delta_{\mathbf{q}+\mathbf{q}'} + S(\mathbf{q}+\mathbf{q}') - 1] \tag{5.209}$$

The prime on the bracket means, in doing the averages, to omit the interaction of a particle with itself. This omission produces the -1 term on the right. The term $N_e \delta_{\mathbf{q}+\mathbf{q}'}$ is the same one included in the above derivation, which gave plasmons. The other term gives $G_H(q)$. It is already proportional to $\rho(\mathbf{q}, t)$, so that $G_H(q)$ is obtained without difficulty. The electron–electron interactions give a term

$$\begin{aligned} &\langle \rho(\mathbf{q}, t)\rangle \frac{1}{\nu} \sum_{\mathbf{q}'} v_{q'} \left(\frac{-\mathbf{q}\cdot\mathbf{q}'}{m} \right) [N_e \delta_{\mathbf{q}+\mathbf{q}'} + S(\mathbf{q}+\mathbf{q}') - 1] \\ &\quad = \omega_p^2 \langle \rho(\mathbf{q}, t)\rangle [1 - G_H(q)] \end{aligned} \tag{5.210}$$

$$G_H(q) = \frac{1}{n_0} \int \frac{d^3q'}{(2\pi)^3} \frac{\mathbf{q}\cdot\mathbf{q}'}{q^{2'}} [S(\mathbf{q}+\mathbf{q}') - 1]$$

$$\omega^2 \rho(\mathbf{q}, t) = \sum_{\mathbf{p}\sigma} (\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}})^2 \langle C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}\rangle + \omega_p^2 \langle \rho(\mathbf{q}, \mathbf{t})\rangle [1 - G_H(q)] \tag{5.211}$$

The factor of $G_H(q)$ provides a modification of the plasmon dispersion at nonzero q due to electron correlations. The formula for $G_H(q)$ is the same one which was announced in (5.198) after changing dummy variables $\mathbf{k} = -\mathbf{q}'$.

The Singwi–Sjölander formula for $G_H(q)$ can be evaluated with only a knowledge of $S(\mathbf{q})$. It, in turn, is obtained from a knowledge of $\varepsilon(\mathbf{q}, \omega)$ through the relationship (5.117) derived earlier:

$$S(\mathbf{q}) = -\frac{1}{n_0 v_q} \int_0^\infty \frac{d\omega}{\pi} \operatorname{Im}\left[\frac{1}{\varepsilon(\mathbf{q}, \omega)}\right]$$

But $\varepsilon(\mathbf{q}, \omega)$ depends on $G_H(q)$. The three equations (5.198), (5.187), and (5.117) form a triad which link the three functions $G_H(q)$, $S(\mathbf{q})$, and $\varepsilon(\mathbf{q}, \omega)$. They are solved self-consistently on the computer, which must be done for each value of r_s, since the results depend on density. In

TABLE 5.2 Constants A and B in the Hubbard local field

r_s	Vashista and Singwi (1972)		Singwi *et al.* (1970)	
	A	B	A	B
1	0.70853	0.36940	0.7756	0.4307
2	0.85509	0.33117	0.8994	0.3401
3	0.97805	0.30440	0.9629	0.2924
4	1.08482	0.28430	0.9959	0.2612
5	1.17987	0.26850	1.0138	0.2377
6	1.26569	0.25561	1.0216	0.2189

discussing some later versions of $G_H(q)$, they (Singwi *et al.*, 1970; Vashishta and Singwi, 1972) remarked that their results could be adequately fitted by the simple expression

$$G_H(q) = A\{1 - \exp[-B(q/q_{TF})^2]\} \tag{5.212}$$

The constants A and B, both dimensionless, are given in Table 5.2 for different values of r_s. Since they depend smoothly on density, the parameters A and B may be obtained for other values of r_s by interpolation. This form for $G_H(q)$ fits their computed one well at small and intermediate values of q/k_F but not at larger values. However, G_H is relatively uninportant at large q, so this drawback is not serious. In the next section a number of properties of the electron gas are calculated using all four dielectric functions. In general the accuracy increases in the order they were present: Thomas–Fermi is least accurate, while Singwi–Sjölander is the best.

5.5.5. Local Field Corrections

The Hubbard local field correction is actually a function of frequency $G_H(\mathbf{q}, \omega)$. The frequency dependence is much less well known than the dependence upon wave vector. Numerous theorists (Niklasson, 1974; Santoro and Giuliani 1988; Vignale, 1988) have derived the exact limit of $G_H(\mathbf{q}, \omega)$ and $G_S(\mathbf{q}, \omega)$ in the limit of large q. They are shown in Table 5.3. The local field for spin correlations $G_S(\mathbf{q}, \omega)$ is introduced later in the chapter. In taking these limits, the result depends upon the behavior of ω while $q \to \infty$. The case for $\omega = 0$ and $\hbar\omega = \varepsilon_q$ give the same limit in two (2D) and three (3D) dimensions. The case that $\omega \neq 0$, but not diverging, give different results in 2D and 3D. Here $g(0)$ is the pair distribution function at $\mathbf{r} = 0$. The usual method of deriving these results is to derive equations for the

TABLE 5.3 Exact asymptotic limits ($q \to \infty$) of the local fields for charge $G_H(q, \omega)$ and spin $G_S(q, \omega)$. The first row is for $\omega = 0$ the second row is for nonzero ω. The third row scales $\hbar\omega$ with ε_q. From Santoro and Giuliani (1988, used with permission)

	3D		2D	
	G_H	G_S	G_H	G_S
$\omega = 0$	$1 - g(0)$	$g(0)$	$1 - g(0)$	$g(0)$
$\omega \neq 0$	$\frac{2}{3}[1 - g(0)]$	$\frac{1}{3}[4g(0) - 1]$	$1 - g(0)$	$g(0)$
$\hbar\omega = \varepsilon_q$	$1 - g(0)$	$g(0)$	$1 - g(0)$	$g(0)$

response functions using the equations of motion. The correction terms vanish in the limit of large wave vector and frequency, so that it is possible to obtain exact limits. They find at large q, and for $\omega \neq 0$, the asymptotic limits

$$G_H(\mathbf{q}, \omega) = \frac{1}{2N_e} \sum_{\mathbf{q}'} \sum_{\sigma\sigma'} \left[\frac{(\mathbf{q} \cdot \mathbf{q}')^2 v(\mathbf{q}')}{q^4 v(\mathbf{q})} - 1 \right] [S_{\sigma\sigma'}(\mathbf{q}') - \delta_{\sigma\sigma'}]$$
$$G_S(\mathbf{q}, \omega) = \frac{1}{2N_e} \sum_{\mathbf{q}'} \sum_{\sigma\sigma'} \left[\frac{(\mathbf{q} \cdot \mathbf{q}')^2 v(\mathbf{q}')}{q^4 v(\mathbf{q})} - \mathrm{sgn}(\sigma\sigma') \right] [S_{\sigma\sigma'}(\mathbf{q}') - \delta_{\sigma\sigma'}] \tag{5.213}$$

$$S_{\sigma\sigma'}(\mathbf{q}) = \frac{2}{N_e} \langle \rho_\sigma(-\mathbf{q}) \rho_{\sigma'}(\mathbf{q}) \rangle - \frac{N_e}{2} \delta_{\mathbf{q}=0} \tag{5.214}$$

$$\rho_\sigma = \sum_{\mathbf{p}} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p},\sigma} \tag{5.215}$$

These formulas have the flavor of the Singwi–Sjölander theory, which is not surprising since they were derived in a similar fashion. In practice, the local field corrections are rarely needed in these limits of large variables. However, exact results are always useful as benchmarks for approximate theories.

Another method to obtain information regarding the local field corrections is from experiments. Larson *et al.* (1996) have determined the local field correction $G_H(q)$ in metallic aluminum from an analysis of x-ray scattering. The scattering measures $S(\mathbf{q}, \omega)$ for a fixed value of $\mathbf{q}$ as a function of ω. This data is compared to the theoretical result which is written in the form

$$S(\mathbf{q}, \omega) \propto -\mathrm{Im} \left[\frac{1}{1 - v_q[1 - G_H(\mathbf{q})] P^{(1)}(\mathbf{q}, \omega)} \right] \tag{5.216}$$

The RPA polarization $P^{(1)}(\mathbf{q}, \omega)$ was calculated by Fleszar *et al.* (1995) using the accurate numerical wave functions for the energy bands of aluminum. For each value of $\mathbf{q}$, the quantity $G_H(\mathbf{q})$ is treated as an adjustable constant to improve the agreement between the computed and experimental values. Repeating the experiment for many different points $\mathbf{q}$ gives a plot of $G_H(\mathbf{q})$ vs. $\mathbf{q}$, which is shown in Fig. 5.14. There is reasonable agreement between the measured values, and the predictions for the jellium model.

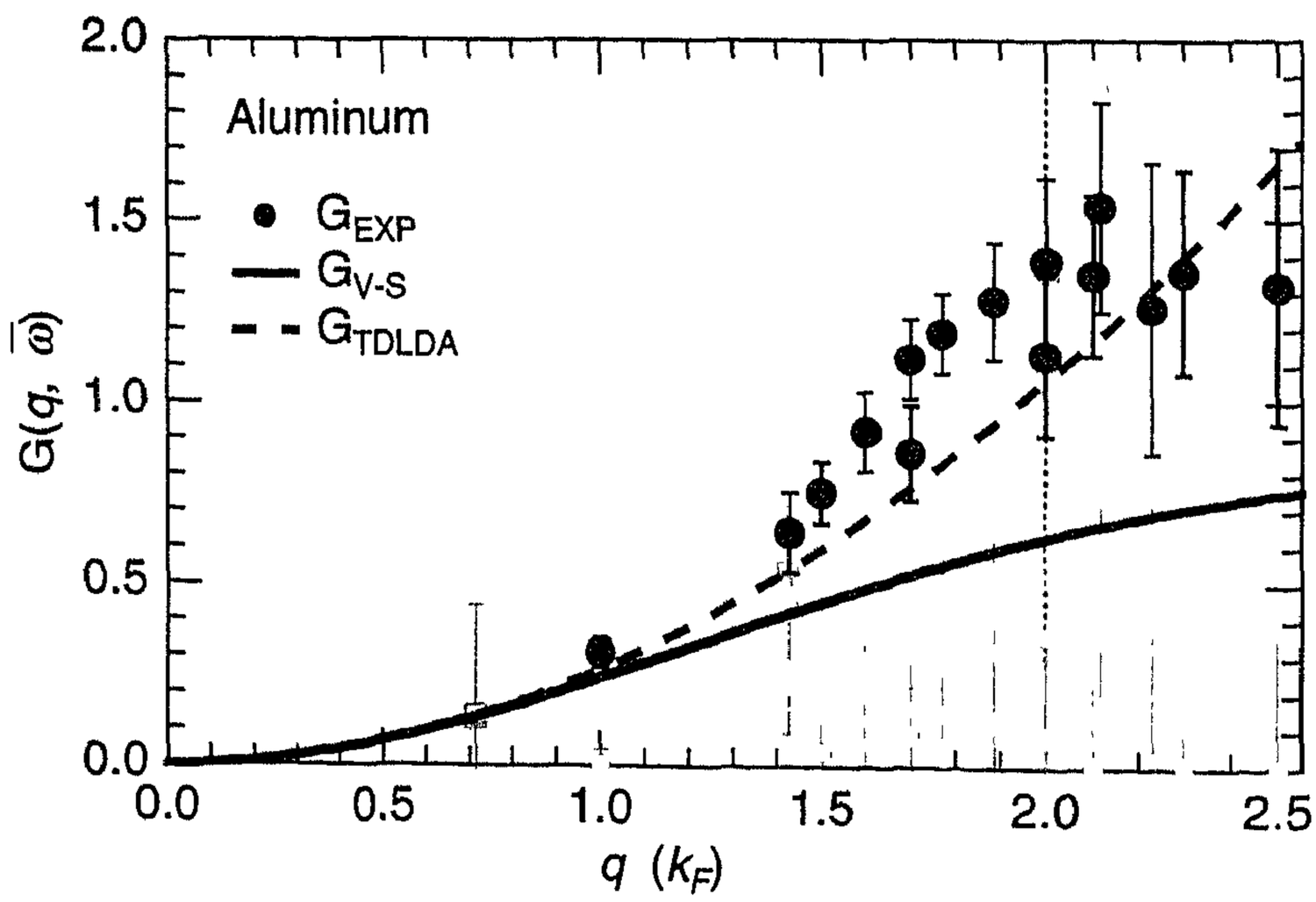


FIGURE 5.14 Measured local field function $G_H(q)$ in metallic aluminum by Larson *et al.* (1996). Also shown are several theoretical curves based upon the homogeneous electron gas (used with permission).

5.5.6. Vertex Corrections

The Hubbard local field factor $G_H(q)$ was derived as a vertex correction to the polarization diagram of the electron gas. This vertex correction is denoted as $\Gamma_H(q)$, and the Hubbard form of the dielectric function is written as

$$\varepsilon_H(q) = 1 - v_q P^{(1)}(q)\Gamma_H(q) \tag{5.217}$$

$$\Gamma_H(q) = \frac{1}{1 + v_q G_H(q) P^{(1)}(q)} \tag{5.218}$$

Where else might vertex corrections enter into the calculations of the electron gas?

The first point is to distinguish between the interaction between two impurity charges $Z_{1,2}$ in the electron gas, compared to the interaction between two electrons. The impurities were assumed not to be electrons. The above screening function is appropriate only for the interaction between two protons in the electron gas, or between other charged ions.

Additional vertex corrections are required when electrons interact among themselves. In some cases one factor is required, while at other times two are required. This confusing situation is now discussed. The first case is the calculation of the screened exchange energy Σ_{sx} of an electron. The self-energy is

$$\Sigma_{sx}(\mathbf{k}, ik_n) = -\frac{1}{\beta}\sum_{iq_m}\int\frac{d^3q}{(2\pi)^3} v_q \frac{\Gamma_H(\mathbf{q}, iq_m)}{\varepsilon(\mathbf{q}, iq_m)} \mathscr{G}^{(0)}(\mathbf{k}+\mathbf{q}, ik_n + iq_m) \tag{5.219}$$

$$\varepsilon(\mathbf{q}, iq_m) = 1 - v_q P^{(1)}(\mathbf{q}, iq_m)\Gamma_H(\mathbf{q}, iq_m) \tag{5.220}$$

$$\Gamma_H(\mathbf{q}, iq_m) = \frac{1}{1 + v_q G_H(q) P^{(1)}(\mathbf{q}, iq_m)} \tag{5.221}$$

$$\frac{\Gamma_H(\mathbf{q}, iq_m)}{\varepsilon(\mathbf{q}, iq_m)} = \frac{1}{1 - v_q[1 - G_H(q)]P^{(1)}(\mathbf{q}, iq_m)} \tag{5.222}$$

The integrand contains a single vertex function $\Gamma_H(\mathbf{q}, iq_m)$. The vertex function should depend upon the k variable. In four-vector notation, it is a function of $\Gamma(k, k+q)$. Neglecting the k dependence is an approximation. All dielectric functions of the Hubbard type, with a $G_H(q)$ factor, make this approximation of neglecting the k dependence of the vertex function.

Figure 5.15 shows the vertex corrections of the screened exchange energy. The wiggly line denotes the interaction $W(q) = v_q/\varepsilon(q)$, where $\varepsilon(q)$ is the Hubbard dielectric function. The first diagram has a single wiggly line and corresponds to the interaction $v_q/\varepsilon(q)$. The second diagram has two wiggly lines. Because of the symmetry, the diagram could be a vertex correction to either end point. Since the diagram occurs only once, it cannot be a vertex correction to both end points. Keeping only the ladder diagrams, then the vertex corrections to the first end point are shown by several more terms in the series. The summation of these terms, the vertex correction, is shown in Feynman diagrams as a shaded triangle.

The key point is that only one end of the large wiggly interaction has a vertex correction. Do not put a vertex correction at each end, since that is overcounting. In Fig. 5.15(b) is shown two Feynman diagrams. The first has a vertex correction at each end of the large wiggly line. This term seems to suggest a correction for each end. However, the second Feynman diagram in Fig. 5.15(b) is topologically identical to the first. The only difference is that a different wiggly line has been drawn "large". That diagram can be considered to be a vertex correction to the first end point, where one of the ladder diagrams itself has a vertex correction. In fact, the correct way to include all ladder diagrams is to have each ladder also have a vertex

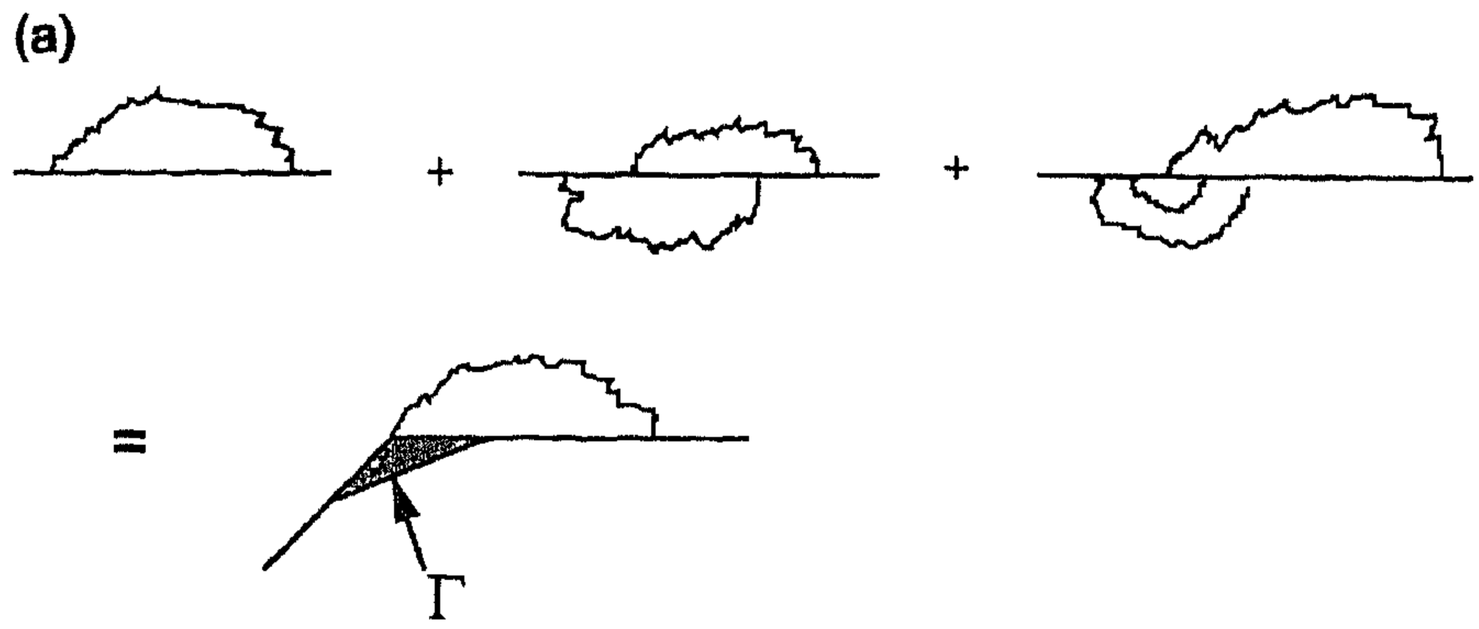


FIGURE 5.15 Vertex corrections to the screened exchange energy.

correction at each end. This set of Feynman diagrams has fractal character (Mahan, 1993). Writing the equations for all ladder corrections makes it obvious that the vertex correction only goes on one end in the screened exchange energy.

There is another way to show the correctness of Eq. (5.219). Ting *et al.* (1975) showed the vertex function Γ_H enters only once in effective interaction. Furthermore, they also showed the same vertex function enters the numerator as enters the dielectric function.

Ting *et al.* (1975) derived (5.219) by starting from the definition of the ground state energy of the interacting electron system

$$E_g = E_{g0} - \frac{\nu}{2\beta}\int_0^1 \frac{d\eta}{\eta}\int \frac{d^3q}{(2\pi)^3}\sum_{iq}\left[\frac{1}{\varepsilon_\eta(\mathbf{q}, iq)} - 1\right] \tag{5.223}$$

$$\varepsilon_\eta(\mathbf{q}, iq) = 1 - \frac{\eta v_q P^{(1)}(\mathbf{q}, iq)}{1 + \eta v_q G_H(q) P^{(1)}(\mathbf{q}, iq)} \tag{5.224}$$

The derivation of this formula starts by defining the ground state energy as the expectation value of the Hamiltonian $\langle H\rangle$. The term E_{g0} is from the kinetic energy. The second term is from the expectation value of the electron–electron interaction $\langle\rho(\mathbf{q})\rho(-\mathbf{q})\rangle$. The latter is related to the inverse dielectric function using (5.112). The coupling constant η has been made dimensionless by having $\eta = 0$ at $e^2 = 0$ and $\eta = 1$ at e^2. $G_H(q)$ may also be a functional of the product $\eta\chi$, where $\chi = v_q P^{(1)}$.

The self-energy can be obtained as the functional derivative of the ground state energy with respect to the occupation number $n_{\mathbf{k}}$. For example, the kinetic energy term can be written as

$$E_{g0} = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} n_{\mathbf{p}\sigma} \tag{5.225}$$

$$\frac{\delta E_{g0}}{\delta n_{\mathbf{p}\sigma}} = \varepsilon_{\mathbf{p}} \tag{5.226}$$

$$\Sigma_{sx}(\mathbf{p}, ip_n) = \frac{\delta E_g}{\delta n_{\mathbf{p}\sigma}} - \varepsilon_{\mathbf{p}} \tag{5.227}$$

The screened exchange energy is the derivative of the energy in (5.223). Assume the only dependence upon $n_{\mathbf{p}\sigma}$ in the integrand is in the susceptibility χ. The functional derivative has the form for $f = 1/\varepsilon(q)$ is

$$\frac{\delta f}{\delta n_{\mathbf{p}\sigma}} = v_q \frac{\delta f}{\delta \chi}\frac{\delta P^{(1)}}{\delta n_{\mathbf{p}\sigma}} \tag{5.228}$$

$$P^{(1)} = \frac{1}{\mathrm{v}}\sum_{\mathbf{p}\sigma} n_{\mathbf{p}\sigma}\left(\frac{1}{\xi_{\mathbf{p}} - \xi_{\mathbf{p}+\mathbf{q}} + iq_n} + \frac{1}{\xi_{\mathbf{p}} - \xi_{\mathbf{p}+\mathbf{q}} - iq_n}\right) \tag{5.229}$$

$$\frac{\delta P^{(1)}}{\delta n_{\mathbf{p}\sigma}} = \frac{1}{\mathrm{v}}[\mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}, \varepsilon_{\mathbf{p}} + iq_n) + \mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}, \varepsilon_{\mathbf{p}} - iq_n)] \tag{5.230}$$

The two Green's function terms in $\delta P^{(1)}/\delta n_{\mathbf{p}\sigma}$ contribute equally, since $\varepsilon(\mathbf{q}, iq_n)$ is a symmetric function of complex frequency iq_n. This identity removes the factor of 2 in front of the interaction term in (5.223). The two steps of coupling constant integral and functional derivative can be arranged to cancel. Introduce the variable $y = \eta\chi$, which becomes the variable of both integration and differentiation:

$$\int_0^1 \frac{d\eta}{\eta}\frac{df(\eta\chi)}{d\chi} = \frac{1}{\chi}\int_0^\chi dy \frac{df(y)}{dy} = \frac{1}{\chi}[f(\chi) - f(0)] \tag{5.231}$$

When $f(\chi) = 1/\varepsilon = 1/(1 - \Gamma\chi)$ then $f(\chi) - f(0) = \chi\Gamma/\varepsilon$. Collecting all of these results produces (5.219). This derivation is independent of the particular form for the vertex function Γ. It is also valid for any form for $G_H(q)$ as long as one assumes that $G_H(q)$ is a functional of the susceptibility χ. This derivation makes it clear that the screened exchange energy has only a single factor of the vertex function in its effective interaction.

There are occasions where an effective Coulomb interaction between two electrons has a double power (Γ_H^2) of the vertex correction. The ladder diagrams of the polarization diagrams could themselves have vertex corrections at each end of the wiggly line. The wiggly line is still the interaction $v_q/\varepsilon(q)$. An equation for the vertex correction in this case is, using a four vector notation

$$\begin{aligned}\Gamma(p, p+q) =& 1 - \sum_{q'}\frac{v_{q'}}{\varepsilon(q')}\Gamma(p+q', p+q+q')\Gamma(p, p+q') \\ &\times \Gamma(p+q, p+q+q')\mathscr{G}^{(0)}(p+q')\mathscr{G}^{(0)}(p+q+q')\end{aligned} \tag{5.232}$$

In the approximation of taking $\Gamma(p, k) \approx \Gamma(p - k)$ then this equation simplifies to

$$\Gamma(q) = 1 - \Gamma(q)\sum_{q'}\frac{v_{q'}\Gamma(q')^2}{\varepsilon(q')}\mathscr{G}^{(0)}(p+q')\mathscr{G}^{(0)}(p+q+q') \tag{5.233}$$

The effective interaction for the ladder diagram has a double vertex correction. A double term always happens when two different electrons are interacting. A single vertex occurs when one electron interacts with itself through the average polarization of the media.

The effective interaction in the screened exchange energy is given in Eq. (5.222). The Coulomb interaction v_q is multiplied by the factor of $[1 - G_H(q)]$. This expression is often written down as the effective dielectric function. In writing it this way, a vertex correction has been included. If $G_H(q) = 1$ for some value of $\mathbf{q}$ then the Coulomb interaction is effectively unscreened. In the x-ray scattering from aluminum, that occurs at $q = 1.65k_F$ (Larson *et al.*, 1996).

5.6. PROPERTIES OF THE ELECTRON GAS

There are a number of properties of the electron gas which may be predicted with a knowledge of the dielectric function $\varepsilon(\mathbf{q}, \omega)$. These properties may also be measured for metals. However, one should be careful in applying predictions of the homogeneous electron gas to real metals, because the ion cores often have a significant influence. A favorable comparison to experiment may not, in his case, be the signature of the best theory. Four properties are discussed: the pair distribution function, the screening charge, the compressibility, and the correlation energy.

5.6.1. Pair Distribution Function

The pair distribution function $g(r)$ may be obtained from a knowledge of the static structure factor $S(q)$ through the transform relation (1.445):

$$g(r) = 1 + \frac{3}{2rk_F^3}\int_0^\infty kdk \sin(kr)[S(k) - 1] \tag{5.234}$$

The static structure factor $S(k)$ is obtained from the dielectric function through the relation (5.117). The Thomas–Fermi theory is a static model of dielectric screening, so that it has no prediction regarding $g(r)$. So $g(r)$ is calculated using the theories of RPA, Hubbard, and Singwi–Sjölander. All results are obtained only after a numerical computation.

The three results for $g(r)$ are shown in Fig. 5.16 for six values of r_s. The RPA and Hubbard theories are negative at small values of r. This behavior is viewed as a great deficiency, since $g(r)$ is, by definition, strictly a positive function. The pair distribution function is defined as the probability that another electron is a distance r away from the first; as a probability, it must be positive. The negative values predicted by the RPA and Hubbard theories are regarded as significant failures of these theories. The Singwi–Sjölander theory is slightly negative at large r_s but otherwise positive, so it is much better in this respect. For most metallic densities, it predicts $g(r)$ to be positive for all r. The Singwi–Sjölander theory is a significant improvement over the earlier theories.

Kimball (1973) proved that $g'(0) = g(0)/a_0$ where a_0 is the Bohr radius and $g' = dg/dr$. His proof is based upon the assumption that when two electrons become very close together, the other electrons can be ignored and it becomes a two-body problem. The relative wave function is similar to the hydrogen atom, but with a repulsive interaction. The same result was obtained more rigorously (i.e., including the other electrons) by Vignale (1988).

5.6.2. Screening Charge

The screened potential $V(r)$ about a point charge $Ze\delta(\mathbf{r})$ may be calculated and plotted. It is slightly inconvenient, since $V(r)$ diverges as r^{-1} at small r, as do all Coulomb potentials. Instead, it is customary to plot the density of screening charge $n_s(r)$ about the point impurity. This quantity is nonzero at the origin $r = 0$ and also has the Friedel oscillations at large distance. In the linear screening model, derived in Sec. 5.4, the potential energy from an impurity charge of Ze is

$$V(r) = -Z\int \frac{d^3q}{(2\pi)^3}\frac{v_q}{\varepsilon(q)}e^{i\mathbf{q}\cdot\mathbf{r}} \tag{5.235}$$

Operate on both sides of this equation with ∇^2. From Poisson's equation, $\nabla^2 V$ is proportional to the excess charge density in the system. There is charge from the central impurity $Ze\delta(\mathbf{r})$ and the screening particle density $n_s(r)$. The equation for this screening charge is:

$$\nabla^2 V = 4\pi e^2[Z\delta(\mathbf{r}) - n_s(r)] = 4\pi e^2 Z \int \frac{d^3q}{(2\pi)^3} \frac{e^{i\mathbf{q}\cdot\mathbf{r}}}{\varepsilon(q)} \tag{5.236}$$

$$n_s(r) = Z \int \frac{d^3q}{(2\pi)^3} e^{i\mathbf{q}\cdot\mathbf{r}} \left[1 - \frac{1}{\varepsilon(q)}\right] \tag{5.237}$$

The distribution is spherically symmetric in the homogeneous electron gas, so the angular integrals may be performed at once:

$$n_s(r) = \frac{Z}{2\pi^2 r} \int_0^\infty q dp \sin(qr) \left[1 - \frac{1}{\varepsilon(q)}\right] \tag{5.238}$$

In this case a knowledge of the static dielectric function $\varepsilon(q)$ is sufficient to determine the density of screening charge. The Thomas–Fermi model may also be used to make predictions. It was shown in Sec. 5.5.1 that it predicted the screened potential energy to be $Ze^2 \exp(-q_{TF}r)/r$, and the screening particle density is predicted to be $Zq_{TF}^2 \exp(-q_{TF}r)/r$. This result is not plotted in Fig. 5.17. The rest, which are shown in Fig. 5.17, have to be obtained by numerical Fourier transform, which was first done by Langer and Vosko (1960) for the RPA. All curves are very similar, except for Thomas–Fermi, which is the only one to diverge at $r \to 0$. The RPA (1), Hubbard (2), Singwi–Sjölander (3), and Singwi *et al.* (4) results all show Friedel oscillations which are similar. They are lacking in the Thomas–Fermi theory. The Friedel oscillations are real features of impurities in metals, so that the Thomas–Fermi theory is deficient in several respects. However, the other theories predict remarkably similar results, which should not be surprising. Screening is basically a one-body property, and little correlation is evident in one-body amplitudes. The effects of $G_H(q)$ are much more apparent in properties involving two-body correlations.

5.6.3. Correlation Energies

A knowledge of $\varepsilon(\mathbf{q}, \omega)$ is sufficient for predicting the correlation energy of the electron gas. This fact should be expected, since the correlation energy arises from the correlation hole around the electron. The exchange-correlation hole was put into the derivation of $G_H(q)$. Now $\varepsilon(\mathbf{q}, \omega)$ is manipulated to obtain the correlation energy. The discussion follows Singwi *et al.* (1968).

The Coulomb interaction energy of the electron gas may be written as

$$\frac{1}{2\nu} \sum_{\mathbf{q}\neq 0} v_q \rho(\mathbf{q})\rho(-\mathbf{q}) \tag{5.239}$$

This way of writing the interaction allows an electron to interact with itself. These contributions should be omitted. The average value of $\rho(\mathbf{q})\rho(-\mathbf{q})$ in the electron gas is

$$\langle \rho(\mathbf{q})\rho(-\mathbf{q})\rangle' = N_e[N_e\delta_{\mathbf{q}=0} + S(q) - 1] \tag{5.240}$$

The prime on the bracket means one is to omit the particle interacting with itself. It explains why this differs by -1 from the similar average in (5.111). This equation may be derived by

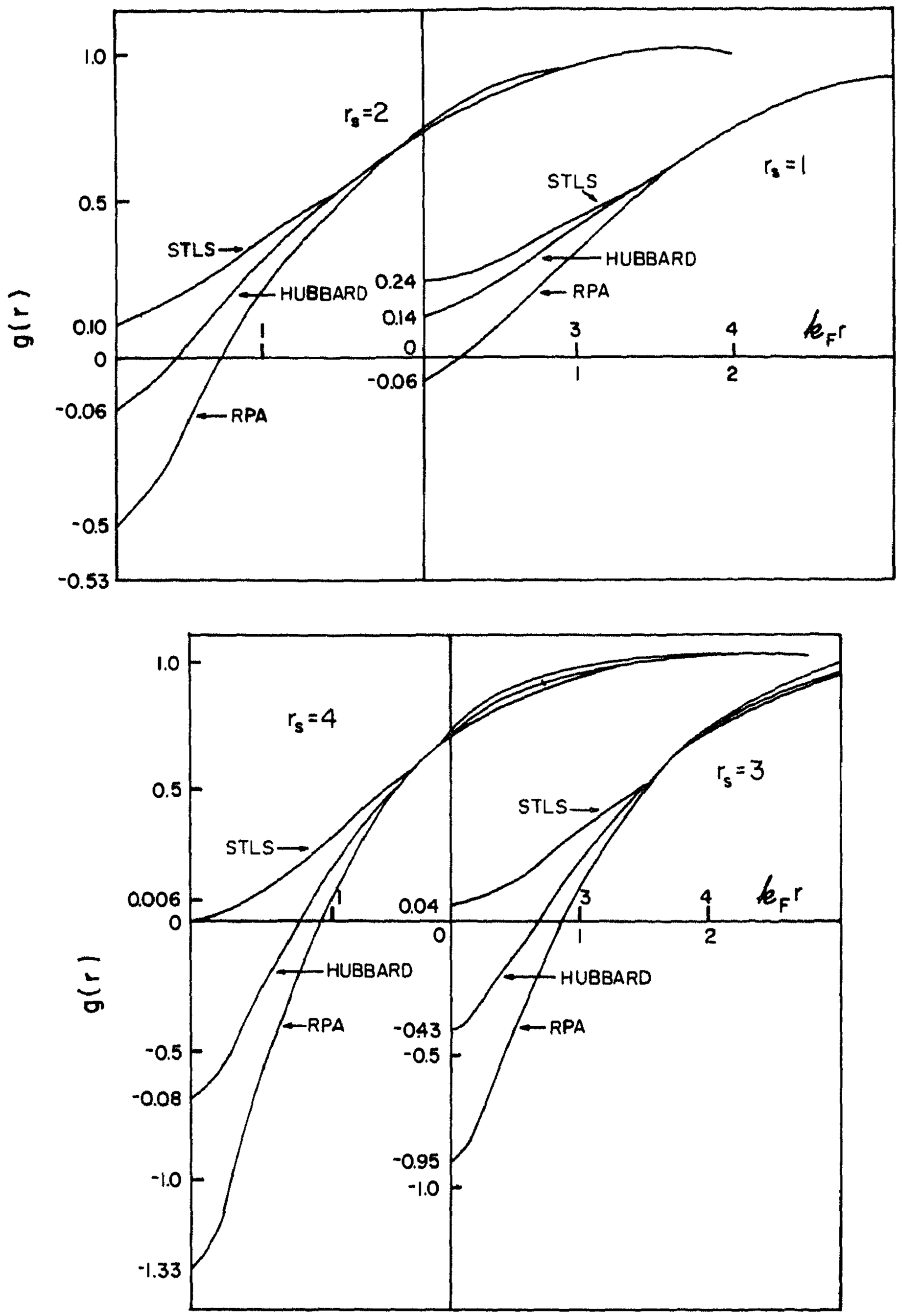


FIGURE 5.16 The pair distribution function $g(r)$ for the electron gas at various densities for several model dielectric functions. *Source*: Singwi *et al.* (1968) (used with permission).

following the same steps used to get (5.209). The interaction energy E_{int} per electron of the electron gas is

$$N_e E_{\text{int}}(e^2) = \frac{N_e}{2\nu} \sum_{\mathbf{q}\neq 0} v_q [S(q) - 1] = 2\pi e^2 N_e \int \frac{d^3q}{(2\pi)^3} \frac{1}{q^2} [S(q) - 1] \tag{5.241}$$

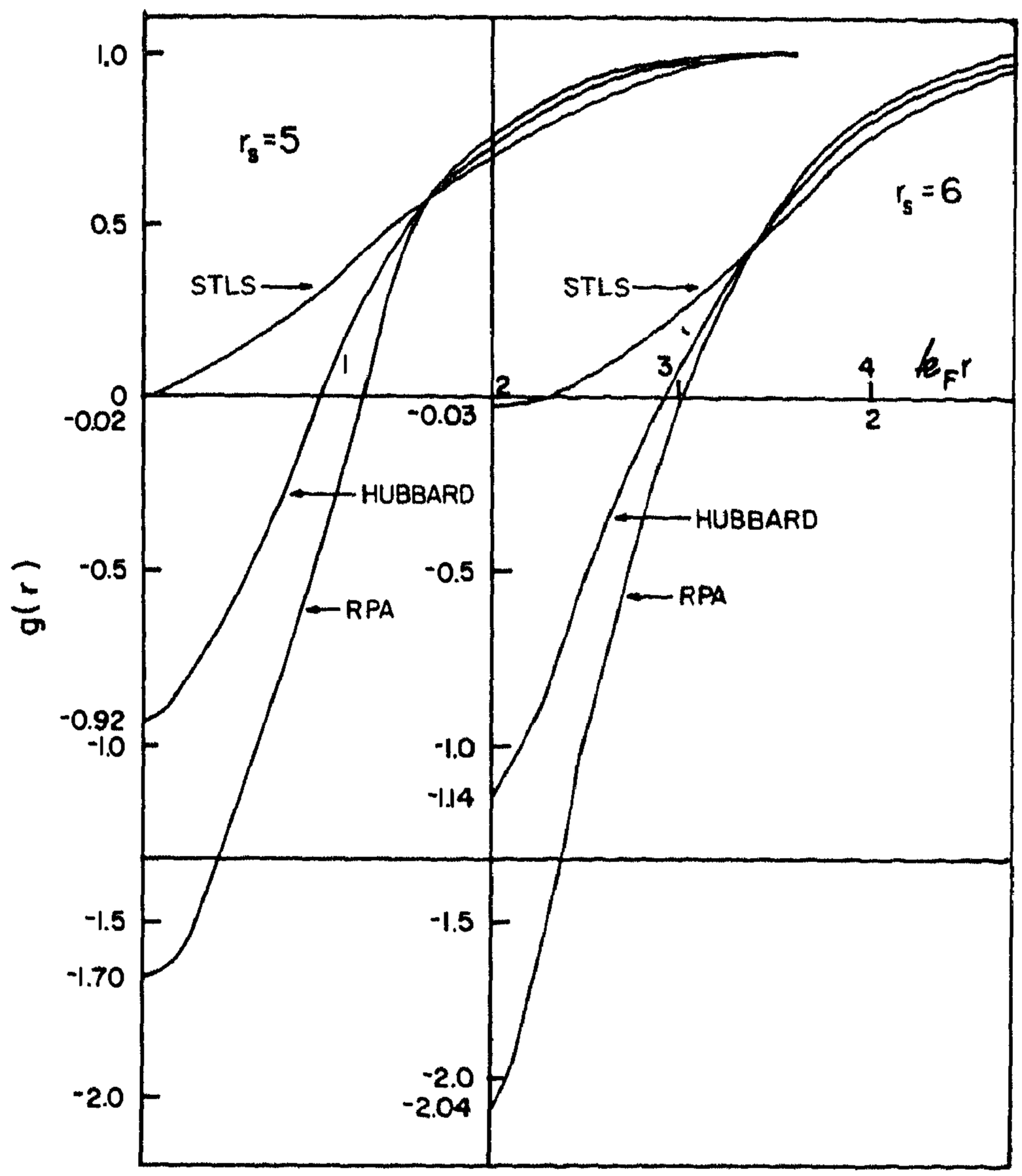


FIGURE 5.16 (*Continued*)

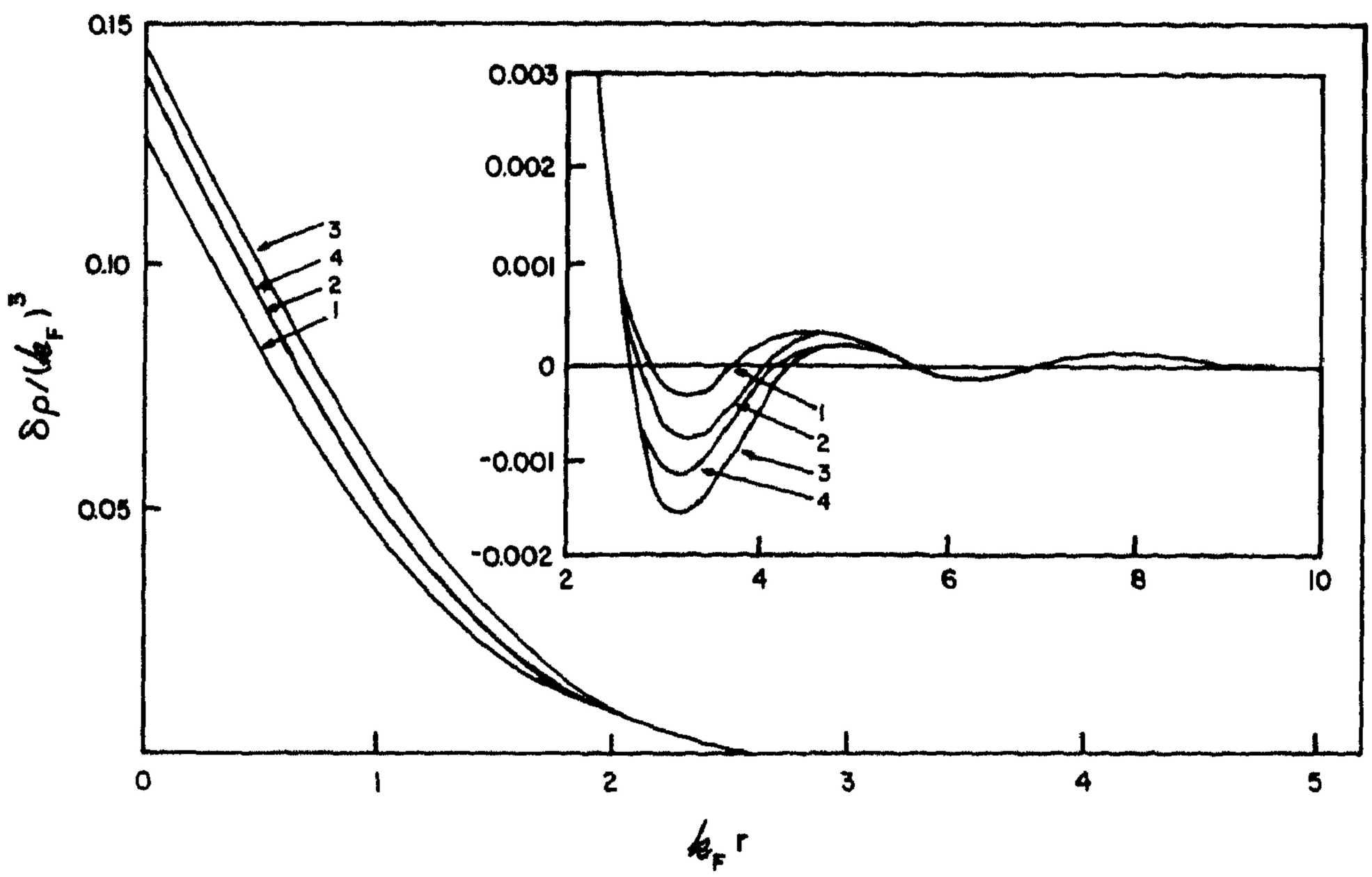


FIGURE 5.17 Screening charge density near a point impurity in the electron gas with $r_s = 3$ according to several model dielectric functions: (1) RPA, (2) Hubbard, (3) Singwi–Sjölander, and (4) Singwi *et al.* *Source*: Singwi *et al.* (1970) (used with permission).

The summation has been converted to an integration over $\mathbf{q}$. In doing this step, remember that the $\mathbf{q} = 0$ term is not in the summation, so that the term $N_e\delta_{\mathbf{q}=0}$ has no effect. The quantity $E_{\text{int}}(e^2)$ is not the Coulomb contribution to the ground state energy. To obtain the ground state energy per particle E_g, one must do a coupling constant integration of the type discussed earlier in Sec. 3.6:

$$E_g = \tfrac{3}{5}E_{F0} + \int_0^{e^2} \frac{d\lambda}{\lambda} E_{\text{int}}(\lambda) \tag{5.242}$$

This equation may be derived starting from (3.299), which is assigned as a problem. Equation (5.242) is an exact result for the ground state energy of the homogeneous electron gas. One must calculate the Coulomb interaction energy for each value of λ up to e^2. In practice, this calculation is the same as finding $S(q)$ as a function of density. The first term $\frac{3}{5}E_{F0}$ is the average kinetic energy, which is calculated assuming that $e^2 = 0$.

The angular integrals in the definition (5.241) of E_{int} may be done, yielding

$$E_{\text{int}}(e^2) = \frac{e^2}{\pi}\int_0^\infty dq[S(q) - 1] \tag{5.243}$$

This expression depends on e^2 through the prefactor on the right. It also has a dependence through $S(q)$, which is decidedly dependent on electron–electron interactions. It is conventional to introduce a dimensionless function

$$\gamma = -\frac{1}{2k_F}\int_0^\infty dq[S(q) - 1] \tag{5.244}$$

which depends on density. In the coupling constant integration (5.242), γ also depends on λ. The interaction energy may be written as

$$E_{\text{int}}(e^2) = -\frac{2}{\pi}e^2 k_F \gamma \tag{5.245}$$

$$= -\frac{4}{\pi r_s}\left(\frac{9\pi}{4}\right)^{1/3}\gamma E_{ry} \tag{5.246}$$

The result has been put into Rydberg units in the second line by using $k_F a_0 = (9\pi/4)^{1/3}/r_s$. Next examine the coupling constant (e^2) dependence of these terms. The Rydberg energy E_{ry} is proportional to e^4 or λ^2. Similarly, $r_s \propto \lambda$ since $r_s \propto a_0^{-1}$ and the Bohr radius is inversely proportional to e^2. Write these dependences as $r_s \to \lambda r_s$ or $e^4 \to \lambda^2 e^4$, where $0 < \lambda < 1$. The interaction energy in this notation is

$$E_{\text{int}}(e^2\lambda) = -\lambda\frac{4}{\pi r_s}\left(\frac{9\pi}{4}\right)^{1/3}\gamma(\lambda r_s) \tag{5.247}$$

The factor γ is also dependent on density, which we write as $\gamma(\lambda r_s)$ or as $\gamma(r_s)$ when $\lambda = 1$. The coupling constant integration may be done, at least formally, and the ground state energy per particle may be obtained from (5.242):

$$E_g = \frac{2.2099}{r_s^2} - \frac{4}{\pi r_s}\left(\frac{9\pi}{4}\right)^{1/3}\int_0^1 d\lambda\gamma(\lambda r_s) \tag{5.248}$$

The coupling constant integration involves only the factor γ. All the other λ's have factored out of the expression. Of course, γ was chosen so for this reason. The correlation energy

consists of all contributions to the ground state energy except kinetic and exchange. Thus the correlation energy is

$$E_c = \frac{0.9163}{r_s} - \frac{4}{\pi r_s}\left(\frac{9\pi}{4}\right)^{1/3}\int_0^1 d\lambda\gamma(\lambda r_s) \tag{5.249}$$

One can verify that this vanishes when $S(q)$ and $\gamma(r_s)$ are calculated in the Hartree–Fock approximation. This expression is evaluated numerically after obtaining $S(q)$ and $\gamma(r_s)$ as functions of density. Some results are given in Fig. 5.18. Four curves are shown, corresponding to RPA, Hubbard, Singwi–Sjölander, and Nozieres–Pines. The last is included to provide a comparison with the earlier plot of the correlation energy in Fig. 5.6. The RPA curve predicts a correlation energy with a magnitude larger than the others. The other three are similar and also similar to that of Wigner. The correlation energy may be obtained with a knowledge of the static structure factor $S(q)$ if it is known for all densities. The Singwi–Sjölander correlation energy is presumed best, because of its success in predicting a positive $g(r)$. The rather direct relationship between $g(r)$ and $S(q)$ implies that the best $g(r)$ is the best $S(q)$, which in turn is the best correlation energy.

The correlation energy is negative. Correlation lowers the energy of the interacting electron gas, and increases its binding. This result is surprising from a naive viewpoint. The main contribution of the correlation energy is to screen the electron–electron interactions. Whereas the Hartree–Fock approximation has a bare Coulomb interaction v_q, the exchange and correlation energy has the screened interaction $v_q/\varepsilon(q, \omega)$. The naive viewpoint is that screening weakens the interaction, which lowers the magnitude of the binding energy. That is, if a Thomas–Fermi dielectric function is employed, then $\varepsilon(q) > 1$ so $v_q/\varepsilon(q) < v_q$ and $|E_{xc}| < |E_x|$. The Thomas–Fermi model for the dielectric function actually predicts a positive correlation energy. The magnitude of the binding energy is reduced. The missing variable is

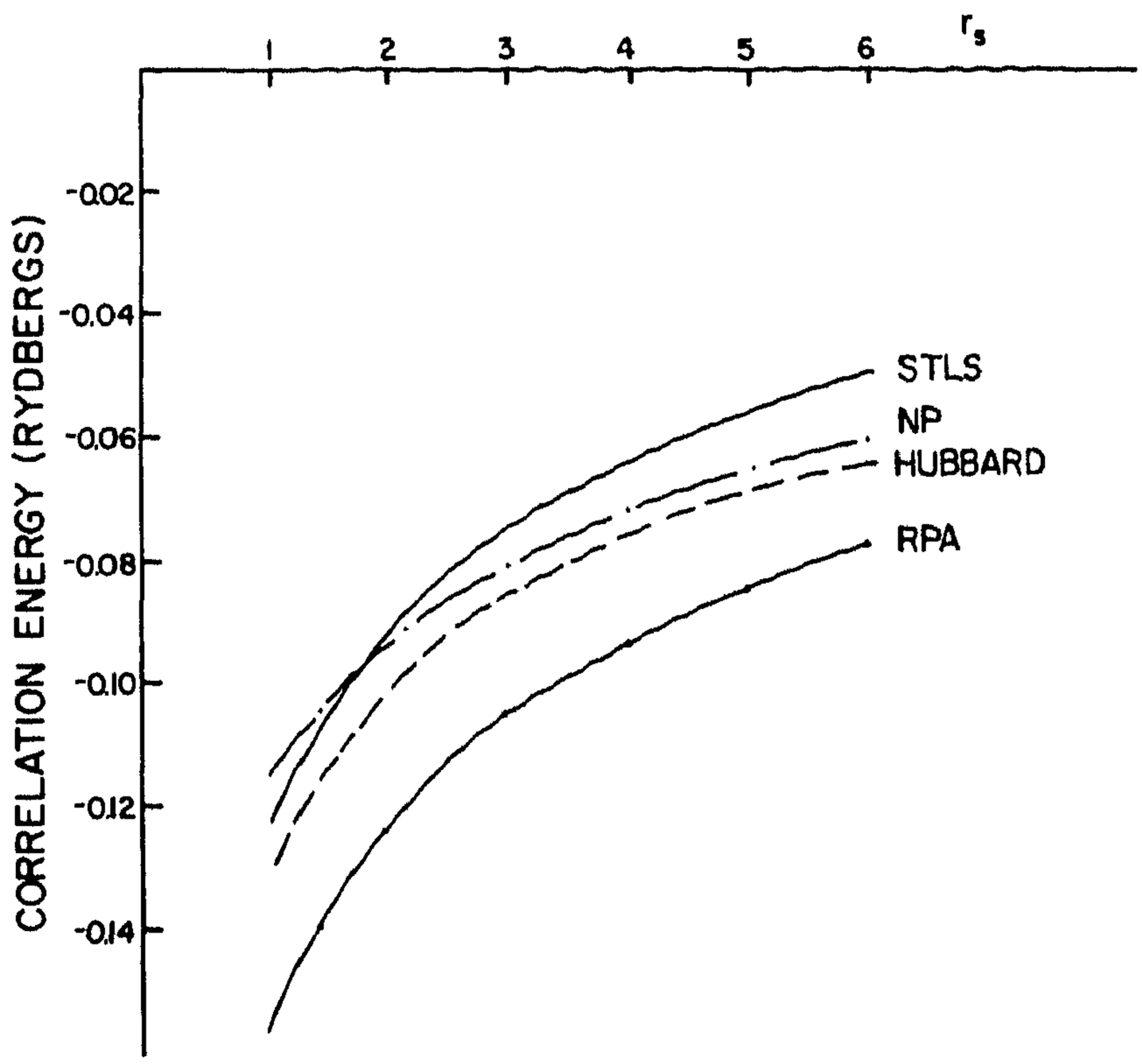


FIGURE 5.18 Correlation energy of the electron gas according to different model dielectric functions: RPA, Hubbard, and Singwi *et al.* Also shown is the interpolation formula of Nozieres and Pines to facilitate comparison with Fig. 5.6.

plasmons. The proper dielectric function has plasmon poles, and these excitations strongly couple to the electron. The interaction with the plasmons provides a significant contribution to the binding energy, and makes the correlation energy negative.

5.6.4. Compressibility

The compressibility of the electron gas is defined as follows. Let E be the total energy of the system, which was calculated in Sec. 5.1. It is an extrinsic quantity; i.e., it is proportional to the number of particles N_e in the system of volume ν. The pressure P is defined as the rate of change in E with volume at constant N_e. The inverse compressibility is the rate of change of P under the same conditions:

$$P = -\left(\frac{dE}{d\nu}\right)_{N_e} \tag{5.250}$$

$$\frac{1}{K} = -\nu\left(\frac{dP}{d\nu}\right)_{N_e} \tag{5.251}$$

The definition of K^{-1} is multiplied by ν so that K is not dependent on the size of the system—it is not dependent on N_e. The compressibility is usually compared to that of a free-particle Fermi gas at zero temperature. There the only energy term is the kinetic energy, which was shown in Sec. 5.1 to be

$$E = \tfrac{3}{5}E_F N_e = \frac{3\hbar^2}{10m}N_e\left(\frac{3\pi^2 N_e}{\nu}\right)^{2/3} \tag{5.252}$$

Differentiate this equation twice with respect to ν, while keeping N_e fixed, and find

$$P = \tfrac{2}{5}E_F n_0 = \frac{\hbar^2}{5m}(3\pi^2)^{2/3}\left(\frac{N_e}{\nu}\right)^{5/3} \tag{5.253}$$

$$\frac{1}{K_f} = \tfrac{2}{3}E_F n_0 = \frac{\hbar^2}{3m}\frac{N_e}{\nu}\left(\frac{3\pi^2 N_e}{\nu}\right)^{2/3} \tag{5.254}$$

The free-particle compressibility is $K_f = 3/(2E_F n_0)$. Most of the numerical results are presented as the ratio K_f/K, where K_f is the above result. An effective mass of unity is assumed.

This procedure for finding the compressibility may be generalized. Ignoring surface effects, the total energy of the system is N_e times the exact ground state energy per particle $E = N_e E_g$. For a fixed N_e, the only volume dependence is through the density dependence of $E_g(n_0)$, where $n_0 = N_e/\nu$. Twice differentiating the total energy yields

$$P = -N_e\frac{dE_g}{d\nu} = n_0^2\frac{dE_g}{dn_0} \tag{5.255}$$

$$\frac{1}{K} = -\nu\left(\frac{dP}{d\nu}\right)_{N_e} = n_0\frac{dP(n_0)}{dn_0} = 2n_0^2E_g' + n_0^3E_g'' \tag{5.256}$$

$$= n_0^2\frac{d^2}{dn_0^2}(n_0E_g) \tag{5.257}$$

The last line shows that the compressibility may also be represented as the second derivative of the quantity $f(n_0) = n_0E_g$. Since $E = N_eE_g = \nu f$, $f(n_0)$ is just the energy per unit volume.

The compressibility is easily found as the second derivative of this quantity multiplied by n_0^2. The calculation of the ground state energy E_g was described in Sec. 5.6.3. Using this method, one may find $f(n_0)$ and its derivatives numerically.

Another relation for the compressibility may be obtained by using the theorem of Seitz, which was given in (5.34):

$$\mu = \frac{df(n_0)}{dn_0} \tag{5.258}$$

$$\frac{1}{K} = n_0^2 \frac{d\mu}{dn_0} \tag{5.259}$$

Here μ is the chemical potential of the electrons, evaluated ignoring surface effects. The compressibility is related to the change in chemical potential with density. In some cases it may be easier to calculate the properties of a particle on the Fermi surface rather than the total energy. In particular, only a single derivative is needed to evaluate (5.259).

As an example of using this formula, the compressibility is calculated in the Hartree–Fock approximation. The starting point is the energy of an electron at the chemical potential

$$\mu = \frac{\hbar^2 k_F^2}{2m} - \frac{e^2 k_F}{\pi} \tag{5.260}$$

It is a quadratic equation for $k_F(\mu)$. Solve it, and then find the electron density from $n_0 = k_F^3/3\pi^2$, where a_0 is the Bohr radius

$$k_F = \frac{1}{\pi a_0} + \sqrt{\frac{1}{(\pi a_0)^2} + \frac{2m\mu}{\hbar^2}} \tag{5.261}$$

$$n_0 = \frac{1}{3\pi^2}\left[\frac{1}{\pi a_0} + \sqrt{\frac{1}{(\pi a_0)^2} + \frac{2m\mu}{\hbar^2}}\right]^3 \tag{5.262}$$

$$\frac{dn_0}{d\mu} = \frac{mk_F^2}{\pi^2\hbar^2}\frac{1}{\sqrt{\frac{1}{(\pi a_0)^2} + \frac{2m\mu}{\hbar^2}}} = \frac{3n_0}{2E_F}\left(\frac{k_F}{\sqrt{\frac{1}{(\pi a_0)^2} + \frac{2m\mu}{\hbar^2}}}\right) \tag{5.263}$$

$$n_0^2 \frac{d\mu}{dn_0} = \frac{2}{3} n_0 E_F \left(\frac{\sqrt{\frac{1}{(\pi a_0)^2} + \frac{2m\mu}{\hbar^2}}}{k_F}\right) = \frac{1}{K_f}\left(1 - \frac{1}{\pi k_F a_0}\right) \tag{5.264}$$

$$\frac{K_f}{K} = 1 - \frac{1}{\pi k_F a_0} = 1 - \frac{r_s}{6.02} \tag{5.265}$$

The Hartree–Fock prediction is that the inverse compressibility is a linear function of r_s, and becomes negative around $r_s = 6.02$. Other predictions are given below.

Another method of evaluating K is through the compressibility sum rule. This rule is an exact relationship between the compressibility and the long-wavelength limit of the static

dielectric function (see Nozieres, 1964). Take the longitudinal dielectric function $\varepsilon(q, \omega)$, set $\omega = 0$, and then take the limit $q \to 0$. In this limit, one obtains

$$\lim_{q\to 0} \varepsilon(q, 0) = 1 + \frac{4\pi e^2}{q^2} n_0^2 K \tag{5.266}$$

By substituting the free-particle compressibility $K_f = 3/(2n_0 E_F)$, the compressibility sum rule may also be written as

$$\lim_{q\to 0} \varepsilon(q, 0) = 1 + \frac{q_{TF}^2}{q^2}\left(\frac{K}{K_f}\right) \tag{5.267}$$

The RPA predicts $K/K_f = 1$, since $\varepsilon_{\text{RPA}} \to 1 + q_{TF}^2/q^2$. It also provides a method of relating K to $G_H(q)$. If one takes the limit $q \to 0$ for the dielectric function (5.187) and remembers that $v_q P^{(1)} \to -q_{TF}^2/q^2$, then

$$\lim_{q\to 0} \varepsilon(q, 0) = 1 - \frac{v_q P^{(1)}}{1 + G_H(q) v_q P^{(1)}} \tag{5.268}$$

$$= 1 + \frac{q_{TF}^2/q^2}{1 - q_{TF}^2 G_H(q)/q^2} \tag{5.269}$$

The function $G_H(q)$ vanishes as $O(q^2)$, so that the limit $G_H(q)/q^2$ yields a constant, which is called a/k_F^2. The constant a is dimensionless. The Hubbard theory predicts that $a = \frac{1}{2}$, so the compressibility is

$$\frac{K_f}{K} = \lim_{q\to 0}\left[1 - q_{TF}^2 \frac{G_H(q)}{q^2}\right] = 1 - a\left(\frac{q_{TF}}{k_F}\right)^2 \tag{5.270}$$

$$\left(\frac{K_f}{K}\right)_{\text{Hubbard}} = 1 - \frac{1}{2}\left(\frac{q_{TF}}{k_F}\right)^2 = 1 - \frac{r_s}{3.01} \tag{5.271}$$

The constant a may also be obtained from the Singwi–Sölander theory, where $G_H(q)$ is given by (5.198):

$$G_H(\mathbf{q}) = -\frac{1}{n_0}\int \frac{d^3k}{(2\pi)^3}\frac{\mathbf{q}\cdot\mathbf{k}}{k^2}[S(\mathbf{q}-\mathbf{k}) - 1] \tag{5.272}$$

Change variables of integration to $\mathbf{Q} = \mathbf{q} - \mathbf{k}$, and evaluate the expression in the limit where $q \to 0$:

$$\begin{aligned}\lim_{q\to 0} G_H(q) &= -\frac{3\pi^2}{k_F^3}\frac{1}{(2\pi)^2}\int_0^\infty Q^2 dQ[S(Q) - 1]\int_{-1}^{1} d\nu \frac{q(q - \nu Q)}{Q^2 + q^2 - 2\nu q Q} \\ &= -\frac{3}{4k_F^3}\int_0^\infty Q^2 dQ[S(Q) - 1]\int_{-1}^{1} d\nu\left[\frac{q^2}{Q^2}(1 - 2\nu^2) + O(q^4)\right] \\ &= -\frac{q^2}{2k_F^3}\int_0^\infty dQ[S(Q) - 1]\end{aligned} \tag{5.273}$$

The combination of factors on the right is the same as in the definition of γ in (5.244). The Singwi–Sjölander theory predicts that $a = \gamma$ or that

$$\frac{K_f}{K} = 1 - \gamma\left(\frac{q_{TF}}{k_F}\right)^2 \tag{5.274}$$

The value of a may be deduced from Table 5.2, since the limit $q \to 0$ yields $a = AB$, where A and B are the constants listed. This value of a varies with density, but it is generally smaller than the Hubbard value of $\frac{1}{2}$ and is nearer to $\frac{1}{4}$.

Two ways have been given to evaluate the compressibility. The first is by finding the ground state energy per particle and differentiating twice. The second is from the long-wavelength limit of the dielectric function. The two methods would give the same result if all quantities could be found exactly. But when using approximate theories for $\varepsilon(q, \omega)$, the two methods of finding the compressibility will give different results. For example, consider the compressibility in the Hartree–Fock approximation. The ground state energy now includes the exchange term, and a second derivative of the ground state energy gives $a = \frac{1}{4}$ as derived earlier. But when the exchange hole is used to get $g(r)$, as at the end of Sec. 5.1, and then $S(q)$, and then $G_H(q)$, and finally a, a different result is found: $a = \frac{3}{8}$. The Hartree–Fock approximation predicts different results in the two methods of calculation. The form of $G_H(q)$ suggested by Vashishta and Singwi (1972) and Schneider *et al.* (1970) has the virtue that it gives almost the same compressibility by both methods of calculation. Denoting their function as $G_{vs}(q)$ while $G(q)$ is the function in (5.198), they are related by

$$G_{vs}(q) = \left(1 + \tfrac{2}{3} n_0 \frac{d}{dn_0}\right) G(q) \tag{5.275}$$

The derivative is taken with respect to density. This form is regarded as being particularly accurate for the compressibility.

Figure 5.19 shows some theoretical compressibilities as a function of r_s. The solid lines are results obtained by differentiating the ground state energy. All the theories except Hartree–Fock predict identical compressibilities when evaluated in this fashion. And the Hartree–Fock result is not very different. Presumably the curve labeled RPA, SS (for Singwi–Sjölander),

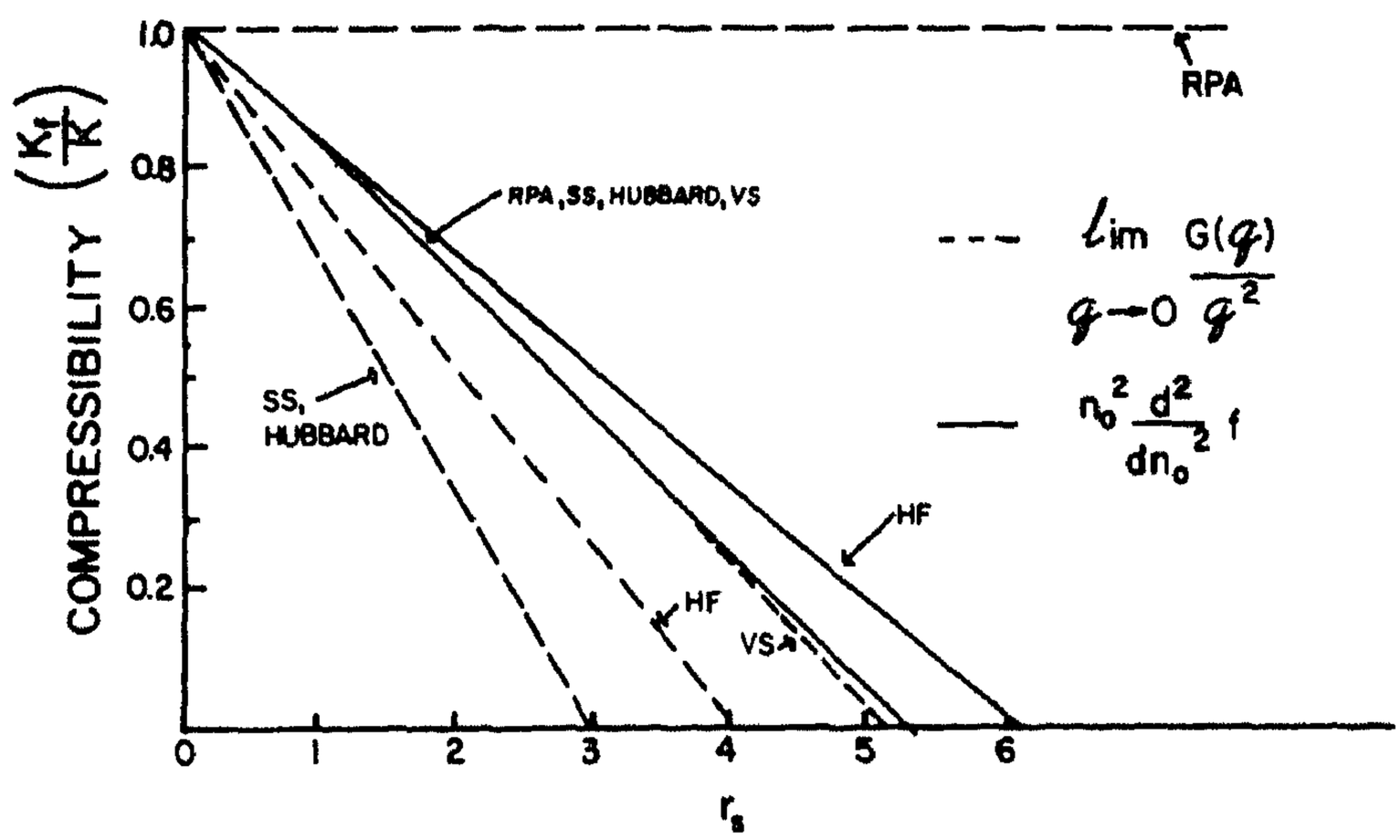


FIGURE 5.19 The compressibility K of the electron gas compared to that of the noninteracting gas K_f. The five theories shown are Hartree–Fock, RPA, Singwi–Sjölander, Hubbard, and Vashishta–Singwi. Each theory makes two predictions: the solid line from the derivative of the ground state energy and the dashed line from the compressibility sum rule.

Hubbard, or VS (for Vashishta–Singwi) is the proper result, since it is given by so many theories.

The dashed lines are the results obtained from the compressibility sum rule: from the limit of $\varepsilon(q)$ or $G_H(q)$ as $q \to 0$. These vary widely among the theories, except for the coincidence of Hubbard with the original Singwi–Sjölander result (see Singwi *et al.*, 1968). The Vashishta–Singwi curve is nearly identical to that calculated by the other method.

The compressibility is predicted to go negative about $r_s = 5$. This result suggests that the electron gas is unstable at densities lower than this critical value. Real metals exist with larger values than $r_s = 5$; i.e., Cs has $r_s = 5.63$, but they are not a homogeneous electron gas. However, metals do not exist with r_s values greater than Cs, which implies that there may be a fundamental limit beyond which the density may no longer be reduced. Unfortunately, this point cannot be tested in three dimensions.

The two-dimensional electron gas has a negative compressibility for r_s above a value of 2.2–2.3. This prediction can be tested. A quasi-two-dimensional gas can be created in a field-effect transistor. The electron density in the surface layer can be controlled by a gate voltage. In silicon devices with very high mobility, it has been found that there is a critical density of $n_c = 0.845 \times 10^{11}\text{cm}^{-2}$ (Kravchenko *et al.*, 1994, 1995). Above this density the layer is metallic, while below this density the layer is insulating. The properties of the insulating state are still under investigation.

5.6.5. Pauli Paramagnetic Susceptibility

The Pauli susceptibility is another one of the many parameters of the electron gas which are calculated using Green's functions. The susceptibility of a free-electron metal such as sodium or aluminum has a number of contributions. One is from the ion cores, and another is the orbital diamagnetism, which is called Landau susceptibility. These two effects must be subtracted from the experimental value in order to deduce the Pauli part, which is due to the spin of the electron.

One way to calculate the Pauli susceptibility is by the use of the Kubo formula derived in Sec. 3.9.1,

$$\chi_{\alpha\beta}(\mathbf{q}, i\omega) = \int_0^\beta d\tau e^{-\omega_n \tau} \langle T_\tau M_\alpha(\mathbf{q}, \tau) M_\beta(-\mathbf{q}, 0)\rangle \tag{5.276}$$

$$M_\alpha(\mathbf{q}) = \mu_0 \sum_{\mathbf{p}\sigma\sigma'} \sigma^{(\alpha)}_{\sigma\sigma'} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma'} \tag{5.277}$$

where $\sigma^{(\alpha)}$ are the Pauli spin matrices. Up or down spin are signified as $\sigma = \pm 1$ for the $s = \frac{1}{2}$ system. Evaluate the static susceptibility, which has $i\omega = 0$. The Hamiltonian is the homogeneous electron gas. Consider the type of terms which occur when expanding the S matrix. The first term, which has no interactions, was evaluated in Sec. 3.9.1. It gave the result for the free-electron gas. which is called $\chi_F = \mu_0^2 N_F$. The first interaction term is of the form

$$\lim_{\mathbf{q}\to 0} \chi_{zz}(\mathbf{q}) = \chi_F - \frac{1}{2\nu} \sum_{\mathbf{pkq}'} \sum_{\sigma\sigma'} v_{q'} \int_0^\beta d\tau \int_0^\beta d\tau_1$$
$$\times \langle T_\tau M_z(\mathbf{q}, \tau) C^\dagger_{\mathbf{p}+\mathbf{q}',\sigma}(\tau_1) C^\dagger_{\mathbf{k}-\mathbf{q}',\sigma'}(\tau_1) C_{\mathbf{k}\sigma'}(\tau_1) C_{\mathbf{p}\sigma}(\tau_1) M_z(-\mathbf{q}, 0)\rangle \tag{5.278}$$

$$M_z(\mathbf{q}) = \mu_0 \sum_{\mathbf{p}\sigma} \sigma C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \tag{5.279}$$

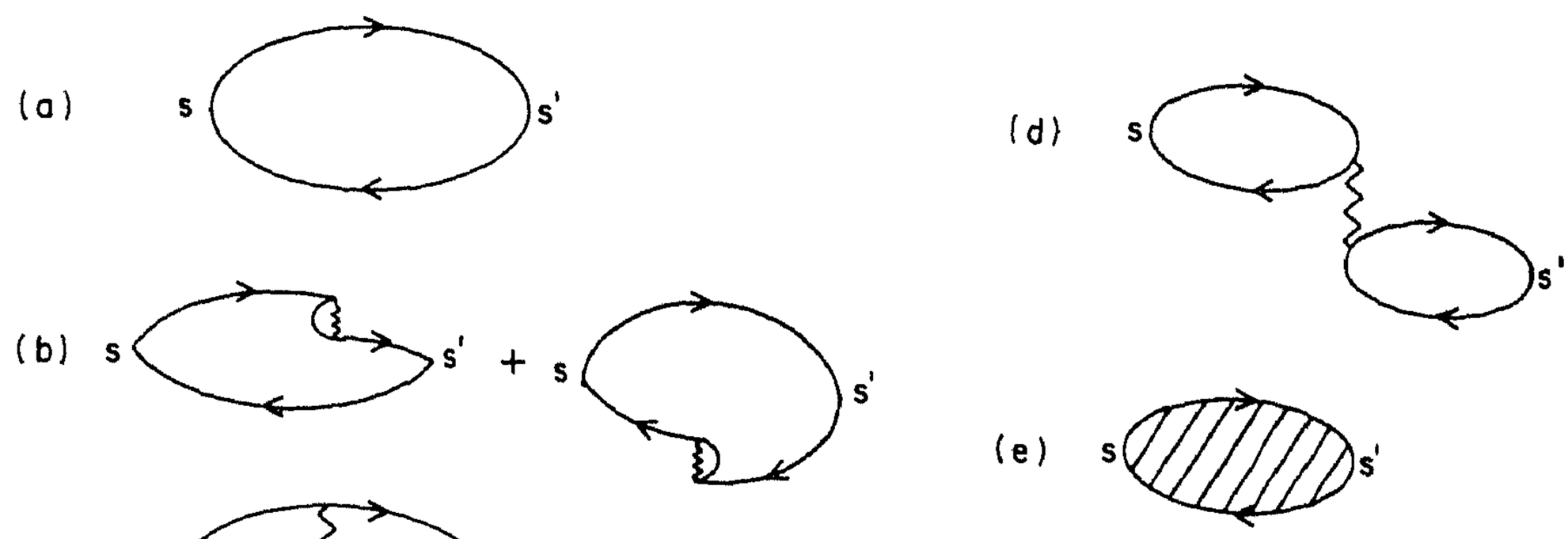


FIGURE 5.20

The several connected terms which result from this contribution are shown in Fig. 5.20. The free-electron term is given by the bubble diagram shown in Fig. 5.20(a). The vertices are labeled s to indicate that they contain a spin operator and so are not the same vertex as found for the density or current operators. Figure 5.20(b) contains the exchange self-energy diagrams for the electron Green's function. These can be incorporated by using $\mathscr{G}$ instead of $\mathscr{G}^{(0)}$ for this propagator, where $\mathscr{G}$ contains the self-energy terms of the electron gas. Figure 5.20(c) has a Coulomb line connecting the two propagators. It is called a vertex correction. It is actually a type of exchange scattering, which contributes to the local field correction $G_S(q)$ of Hubbard and Singwi–Sjölander. Figure 5.20(d) has two closed loops, connected by a Coulomb line. This diagram is zero. It vanishes because each bubble has the correlation function of an M_α operator and a density operator

$$\langle T_\tau M_\alpha(\mathbf{q}, \tau)\rho(-\mathbf{q}, 0)\rangle = 0 \tag{5.280}$$

This combination is zero because of the averaging over $\sigma = \pm 1$. All diagrams with more than one bubble are zero, because the Coulomb interaction is spin independent. Therefore the correlation function is limited to terms with a single bubble, which has an s vertex on each end. The sum of all such contributions is indicated schematically in Fig. 5.20(e). This bubble diagram is called $\bar{P}(\mathbf{q}, \omega)$. It is not the same as the exact $P(\mathbf{q}, \omega)$ which enters into the definition of the longitudinal dielectric function, since the latter may have contributions from more than one bubble if connected by two or more Coulomb lines.

One possible approximation for $\bar{P}$ is to use the Hubbard or Singwi–Sjölander theories:

$$\bar{P}(q, i\omega) = \frac{\mu_0^2 P^{(1)}(q, i\omega)}{1 + v_q G_S(q) P^{(1)}(q, i\omega)} \tag{5.281}$$

As discussed in Sec. 5.5, the local field correction for spin $G_S(q)$ is different than the one $G_H(q)$ for charge fluctuations. For the static susceptibility $\chi_{zz} \equiv \chi$, take the limit $i\omega = 0$ and then the limit $\mathbf{q} \to 0$. In this limit $P^{(1)}$ goes to minus the density of states $N_F = mk_F/\pi^2$, $\chi_F \equiv \mu_0^2 N_F$:

$$\chi = -\mu_0^2 \lim_{q\to 0} \bar{P} = \frac{\mu_0^2 N_F}{1 - v_q G_S(q) N_F} = \frac{\chi_F}{1 - a_S(q_{TF}/k_F)^2} \tag{5.282}$$

In this approximation, the ratio χ/χ_F is similar to the ratio of the compressibilities K/K_f which was found in Sec. 5.6.4. However, the spin parameter a_S is not the same as the charge parameter.

The local field correction for spin $G_S(q)$ has also been extensively investigated. Its exact value is known in the limit of large $(\mathbf{q}, \omega)$ from Santoro and Giuliani (1988), and Vignale (1988). These results are presented in Table 5.3. They were derived in the same fashion as the similar results for charge fluctuations.

5.7. SUM RULES

The longitudinal dielectric function $\varepsilon(q, \omega)$ of the homogeneous electron gas has some exact moments which are useful for checking results. Some of these have already been used and mentioned. For example, the long-wavelength limit is $\lim_{q\to 0} \varepsilon(q, \omega) = 1 + v_q n_0^2 K$, where K is the isothermal compressibility. Similarly, the high-frequency limit is

$$\lim_{\omega\to\infty} \varepsilon(q, \omega) = 1 - \frac{\omega_p^2}{\omega^2} \tag{5.283}$$

where ω_p is the plasma frequency. The frequency ω is large when it is larger than other energies in the system, such as the plasma frequency or the Fermi energy.

Another class of exact results are called sum rules. The "sums" are actually integrals over frequency. Several of them are

$$\int_0^\infty d\omega\omega\varepsilon_2(q, \omega) = \frac{\pi}{2}\omega_p^2 \tag{5.284}$$

$$\lim_{q\to 0}\int_0^\infty \frac{d\omega}{\omega}\varepsilon_2(q, \omega) = \frac{\pi}{2}(v_q n_0 K) \tag{5.285}$$

$$\int_0^\infty d\omega\omega \operatorname{Im}\left[\frac{1}{\varepsilon(q, \omega)}\right] = -\frac{\pi}{2}\omega_q^2 \tag{5.286}$$

$$\lim_{q\to 0}\int_0^\infty \frac{d\omega}{\omega}\operatorname{Im}\left[\frac{1}{\varepsilon(q, \omega)}\right] = -\frac{\pi}{2} \tag{5.287}$$

These last two identities are exact results. The validity of the first two are debated. The last two are rigorous since there is a density–density correlation function for the inverse of the dielectric function, which means that its retarded function is absolutely causal. However, there is no correlation function which defines $\varepsilon(\mathbf{q}, \omega)$. Its retarded function is not required to be causal. However, it is hard to find examples where it is not.

Approximate formulas for $\varepsilon(q, \omega)$ may not automatically obey these relationships. Indeed, approximate formulas are considered virtuous according to how accurately they satisfy these identities. The sign convention has $\varepsilon_2(q, \omega) > 0$ for $\omega > 0$, which implies that $\operatorname{Im}(1/\varepsilon) = -\varepsilon_2/|\varepsilon|^2$ is a negative quantity for $\omega > 0$. The last two sum rules may also be written as

$$\int_0^\infty d\omega\omega\frac{\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2} = \frac{\pi}{2}\omega_p^2 \tag{5.288}$$

$$\lim_{q\to 0}\int_0^\infty \frac{d\omega}{\omega}\frac{\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2} = \frac{\pi}{2} \tag{5.289}$$

These exact results may be proved by a variety of methods. The third relation (5.286) is often called the *longitudinal f-sum rule*. It may be proved from the double commutator

$$C = \langle [[H, \rho(\mathbf{q})], \rho(-\mathbf{q})] \rangle \tag{5.290}$$

where the average is taken, at nonzero temperature, over the thermodynamic states of the system. The Hamiltonian H is for the full homogeneous electron gas. However, the density operator $\rho(\mathbf{q})$ commutes with all terms in this Hamiltonian except the kinetic energy. The first commutator was evaluated in Sec. 5.5.4.

$$\begin{aligned}[H, \rho(\mathbf{q})] &= \sum_{\mathbf{p}\sigma} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}(\varepsilon_{\mathbf{p}+\mathbf{q}} - \varepsilon_{\mathbf{p}}) \\ &= \varepsilon_q \rho(\mathbf{q}) + \frac{1}{m}\sum_{\mathbf{p}\sigma} (\mathbf{q}\cdot\mathbf{p}) C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}\end{aligned} \tag{5.291}$$

The term proportional to $\rho(\mathbf{q})$ will commute with $\rho(-\mathbf{q})$. Only the other term needs to be further evaluated,

$$\begin{aligned}C = \langle [[H, \rho(\mathbf{q})], \rho(-\mathbf{q}] \rangle &= \frac{1}{m}\left\langle \sum_{\mathbf{p}\sigma} \mathbf{q}\cdot\mathbf{p}[C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}, \rho(-\mathbf{q})] \right\rangle \\ &= \frac{1}{m}\sum_{\mathbf{p}\sigma} \mathbf{q}\cdot\mathbf{p}\langle C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}+\mathbf{q},\sigma} - C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma}\rangle \\ &= \frac{1}{m}\sum_{\mathbf{p}\sigma}\langle C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma}\rangle[(\mathbf{p}-\mathbf{q})\cdot\mathbf{q} - \mathbf{p}\cdot\mathbf{q}] = -\frac{q^2}{m}\sum_{\mathbf{p}\sigma} n_{\mathbf{p}} = -\frac{q^2}{m} N_e\end{aligned} \tag{5.293}$$

and yields the simple result that this double commutator is just $-N_e q^2/m$. Next, this double commutator is shown to be proportional to the sum in (5.286). It is evaluated by inserting the complete sets of states $|n\rangle$, $|m\rangle$, which are exact eigenstates of the Hamiltonian, and then collecting terms:

$$\begin{aligned}C = \langle [[H, \rho(\mathbf{q})], \rho(-\mathbf{q})] \rangle &= \sum_{nm}\{e^{-\beta E_n}\langle n|[H, \rho(\mathbf{q})]|m\rangle\langle m|\rho(-\mathbf{q})|n\rangle \\ &\quad - e^{-\beta E_m}\langle m|\rho(-\mathbf{q})|n\rangle\langle|[H, \rho(\mathbf{q})]|m\rangle\} \\ &= \sum_{nm}|\langle n|\rho(\mathbf{q})|m\rangle|^2(e^{-\beta E_n} - e^{-\beta E_m})(E_n - E_m)\end{aligned} \tag{5.294}$$

Earlier in (5.115) it was shown that

$$\mathrm{Im}\left[\frac{1}{\varepsilon(q,\omega)}\right] = -\pi(1 - e^{-\beta\omega})\frac{v_q}{\mathrm{V}}\sum_{nm} e^{-\beta E_n}|\langle n|\rho(\mathbf{q})|m\rangle|^2\delta(\omega + E_n - E_m)$$

so that the sum rule is

$$\int_{-\infty}^{\infty} d\omega\omega\ \mathrm{Im}\left[\frac{1}{\varepsilon(q,\omega)}\right] = -\pi\frac{v_q}{\mathrm{V}}\sum_{nm}|\langle n|\rho(\mathbf{q})|m\rangle|^2(E_n - E_m)(e^{-\beta E_n} - e^{-\beta E_m})$$

Compare the sum rule with the double commutator, and note that they are proportional, which provides the derivation of the sum rule (5.286):

$$\int_{-\infty}^{\infty} d\omega\omega\ \mathrm{Im}\left[\frac{1}{\varepsilon(q,\omega)}\right] = \frac{\pi v_q}{\mathrm{V}} C = -\frac{4\pi e^2 n_0}{m} = -\pi\omega_p^2 \tag{5.295}$$

The integral is from $-\infty$ to ∞, but the negative and positive parts contribute equally, so the left-hand side is twice the integral from 0 to ∞.

The other three sum rules may be derived by using the *Kramers–Kronig relations* (Kronig, 1926; Kramers, 1927). Let $B(\omega)$ be a function of complex variable ω, which is analytic in the upper half plane. Analyticity means that it has no poles or branch cuts above the real axis. It is also assumed that $B(\omega) \to 0$ as $|\omega| \to \infty$.

$$B(\omega) = \frac{1}{i\pi}\int_{-\infty}^{\infty} dt' B(\omega') P\frac{1}{\omega' - \omega} \tag{5.296}$$

The integral is along the real axis. On the real axis, the function $B(\omega)$ has real and imaginary parts, which are called B_1 and B_2, so that $B(\omega) = B_1(\omega) + iB_2(\omega)$. They are related by taking the real and imaginary parts of (5.296):

$$B_1(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} dt' B_2(\omega') P\frac{1}{\omega' - \omega} \tag{5.297}$$

$$B_2(\omega) = -\frac{1}{\pi}\int_{-\infty}^{i} nftydt' B_1(\omega') P\frac{1}{\omega' - \omega} \tag{5.298}$$

These two identities are the Kramers–Kronig relations. They are useful throughout all areas of physics. The present interest is in applying them to the real and imaginary parts of the dielectric function $\varepsilon(q, \omega) = \varepsilon_1 + i\varepsilon_2$. Another application is to define the refractive index $n(\omega)$ and extinction coefficient $k(\omega)$, where $\tilde{n} = n + ik = \sqrt{\varepsilon}$. The complex refractive index $\tilde{n}$ obeys the conditions of the theorem and hence the preceding identities. There are other applications in optics, scattering theory, and virtually all branches of physics. All of these applications, including the present sum rules, assume that the dielectric function is causal so has no poles in the upper half plane of frequency. This assumption is not firmly established for ε but is for $1/\varepsilon$.

The other sum rules are easy to prove as simple applications of the Kramers–Kronig relations. First take the case where $B(\omega) = 1/\varepsilon(q, \omega) - 1$. The -1 term is included to make B vanish as $|\omega| \to \infty$. Using the first Kramers–Kronig relation gives

$$\mathrm{Re}\left[\frac{1}{\varepsilon(q, \omega)}\right] - 1 = \frac{1}{\pi}\int_{-\infty}^{\infty} d\omega' \,\mathrm{Im}\left[\frac{1}{\varepsilon(q, \omega')}\right] P\frac{1}{\omega' - \omega} \tag{5.299}$$

The right-hand side may be simplified by using the fact that $\mathrm{Im}[1/\varepsilon(q, \omega')]$ is asymmetric in frequency. The integrand $(-\infty, 0)$ is evaluated by changing $\omega \to -\omega$ and combining this with the $(0, \infty)$ part to give

$$\mathrm{Re}\left[\frac{1}{\varepsilon(q, \omega)}\right] = 1 + \frac{2}{\pi}\int_{0}^{\infty} d\omega' \mathrm{Im}\left[\frac{1}{\varepsilon(q, \omega')}\right] P\frac{\omega'}{\omega'^2 - \omega^2} \tag{5.300}$$

This identity is also a sum rule but not one of the simple ones in (5.284)–(5.287). However, they are special cases of this result. First take the limit where $\omega \to 0$:

$$\frac{1}{\varepsilon(q)} = 1 + \frac{2}{\pi}\int_{0}^{\infty} \frac{d\omega'}{\omega'} \mathrm{Im}\left[\frac{1}{\varepsilon(q, \omega')}\right] \tag{5.301}$$

The quantity $\varepsilon(q, \omega = 0) \equiv \varepsilon(q)$ is real, so (5.301) expresses the static dielectric function as an integral over the loss function. The sum rule (5.287) is obtained by taking the limit where $q \to 0$. The quantity $\varepsilon(q)$ is evaluated from the compressibility sum rule (5.266),

$$\lim_{q\to 0} \varepsilon(q) = 1 + v_q n_0 K \to O\left(\frac{1}{q^2}\right) \to \infty \tag{5.302}$$

$$\lim_{q\to 0} \frac{1}{\varepsilon(q)} = \lim_{q\to 0} \frac{q^2}{q^2 + 4\pi e^2 n_0 K} = 0 \tag{5.303}$$

which shows that the left-hand side of (5.301) is zero. The terms on the right give

$$\lim_{q\to 0} \int_0^\infty \frac{d\omega'}{\omega'} \operatorname{Im}\left[\frac{1}{\varepsilon(q, \omega')}\right] = -\frac{\pi}{2} \tag{5.304}$$

which is (5.287).

The longitudinal f-sum rule was proved by taking commutators. The remaining two sum rules, (5.284) and (5.285), are obtained from the Kramers–Kronig relations using $\varepsilon - 1$ for $B(\omega)$. Again the 1 term is included so that $\varepsilon - 1$ vanishes at large values of $|\omega|$. From (5.298)

$$\varepsilon_1(q, \omega) - 1 = \frac{1}{\pi} \int_{-\infty}^{\infty} d\omega' \varepsilon_2(q, \omega') P \frac{1}{\omega' - \omega} \tag{5.305}$$

$$\varepsilon_1(q, \omega) = 1 + \frac{2}{\pi} \int_0^\infty d\omega' \varepsilon_2(q, \omega') P \frac{\omega'}{\omega'^2 - \omega^2} \tag{5.306}$$

The second equation here is derived using the fact that $\varepsilon_2(q, \omega)$ is antisymmetric in frequency. This equation is evaluated for $\omega = 0$:

$$\varepsilon_1(q) = 1 + \frac{2}{\pi} \int_0^\infty \frac{d\omega'}{\omega'} \varepsilon_2(q, \omega') \tag{5.307}$$

Equation (5.307) is an exact relation between the imaginary part of the dielectric function $\varepsilon_2(q, \omega)$ and the static part $\varepsilon_1(q)$. In the limit where $q \to 0$, the compressibility sum rule (5.302) gives $\varepsilon_1 = 1 + v_q n_0^2 K$, so that

$$\lim_{q\to 0} \int_0^\infty \frac{d\omega'}{\omega'} \varepsilon_2(q, \omega') = \frac{\pi}{2} v_q n_0^2 K \tag{5.308}$$

which establishes (5.285).

The remaining sum rule (5.284) is found from the limit $\omega \to \infty$ in (5.306):

$$\lim_{\omega\to\infty} \varepsilon_2(q, \omega) = 1 - \frac{2}{\pi\omega^2} \int_0^\infty d\omega' \omega' \varepsilon_2(q, \omega') + O\left(\frac{1}{\omega^4}\right) \tag{5.309}$$

By using (5.283), the left-hand side becomes $1 - \omega_p^2/\omega^2$, and equating terms of order ω^{-2} produces the result

$$\int_0^\infty d\omega' \omega' \varepsilon_2(q, \omega') = \frac{\pi}{2} \omega_p^2 \tag{5.310}$$

All four sums rules have been demonstrated. These are not the only sum rules which exist. Others are sometimes found useful, in particular the third moment (Mihara and Puff, 1968; Hopfield, 1970).

It is curious that the first moment sum rules of both $\varepsilon_2(q, \omega)$ and $\operatorname{Im}[1/\varepsilon(q, \omega)]$ yield the same result $(\pi/2)\omega_p^2$. And this result is independent of q. One should not conclude that ε_2 and

$\mathrm{Im}(1/\varepsilon)$ are identical. Earlier, in Fig. 5.11, it was shown that these two functions are similar at large q but are quite different at small values of q/k_F.

Plasmons play a particularly important role in these sum rules at small values of q. In the limit where $q \to 0$, from (5.170) then

$$\lim_{q\to 0} \varepsilon(q, \omega) = 1 - \frac{\omega_p^2}{\omega^2} + O(\omega^{-4}) \tag{5.311}$$

$$\lim_{q\to 0}\left[\frac{1}{\varepsilon(q, \omega)} - 1\right] = \frac{\omega^2}{\omega^2 - \omega_p^2} - 1 = \frac{\omega_p^2}{\omega^2 - \omega_p^2} \tag{5.312}$$

$$= \frac{\omega_p}{2}\left(\frac{1}{\omega - \omega_p} - \frac{1}{\omega - \omega_p}\right) \tag{5.313}$$

For a retarded function, let $\omega \to \omega + i\delta$, so that the imaginary part of this expression gives

$$\lim_{q\to 0} \mathrm{Im}\left[\frac{1}{\varepsilon(q, \omega)}\right] = -\frac{\pi\omega_p}{2}[\delta(\omega - \omega_p) - \delta(\omega - \omega_p)] \tag{5.314}$$

This relation is rigorously valid only in the limit where $q \to 0$. Nevertheless, it is curious that it satisfies the last two sum rules exactly:

$$\lim_{q\to 0} \int_0^\infty \omega d\omega\, \mathrm{Im}\left[\frac{1}{\varepsilon(q, \omega)}\right] = -\frac{\pi}{2}\omega_p^2 \tag{5.315}$$

$$\lim_{q\to 0} \int_0^\infty \frac{d\omega}{\omega} \mathrm{Im}\left[\frac{1}{\varepsilon(q, \omega)}\right] = -\frac{\pi}{2} \tag{5.316}$$

This result shows that plasmons provide all the contributions to $\mathrm{Im}[1/\varepsilon(q, \omega)]$ in the limit where $q \to 0$. Since the interpretation of $\mathrm{Im}[1/\varepsilon(q, \omega)]$ is that it describes longitudinal excitations of the electron gas, plasmons are the only excitation in the limit where $q \to 0$. Any other excitation would make a contribution to the sum rule, which is impossible since plasmons provide the entire result. Any other excitation has zero strength in the limit where $q \to 0$.

The *single-mode approximation* is to assume the spectral function is given by a single excitation $\omega_p(q)$. The approximation is only accurate in the electron gas at small values of wave vector q, but is often used for all wave vectors. It will be used again in Chapter 11 in describing the excitations in correlated liquids.

5.8. ONE-ELECTRON PROPERTIES

This section discusses the single-particle properties of electrons in the homogeneous electron gas. They include the effective mass and mean free path. The present section is put at the end of this chapter deliberately to make a point: the one-electron properties are not very important for discussions of most features of the electron gas. The discussions of ground state energies and pair correlations did not rely significantly on the properties of the individual electrons. The excitation spectra of the electron gas are also not very dependent on the properties of single electrons. In fact, the converse is true: the collective properties of the electron gas need to be known in order to discuss the behavior of a single electron. This single electron interacts with the other electrons, which are represented as collective excitations. The

order of presentation has been to first describe the collective excitations. It was done assuming that the particles were nearly free (i.e., RPA) or had some correlation (i.e., Hubbard, Singwi–Sjölander). These calculations ignored the finite mean free path, or possible effective mass changes, of the electron. They do not change these collective properties to a significant degree, as long as the mass changes are small, and the mean free path is long.

The original assumption is that electron states in the electron gas are noninteracting plane waves, and the final description is not very different. The self-energy effects do not alter this basic picture.

If the calculations were to show that something drastic happened to the individual electrons—that they were magnetic or superconducting or semiconducting—obviously the collective properties would be altered in a significant fashion. It is important that the initial eigenstates are an approximate description of the actual physical system. It is only in this circumstance that one finds that the one-particle self-energies have little influence on collective properties.

The one-electron properties are obtained from a calculation of the electron self-energy due to electron–electron interactions. This self-energy was discussed earlier, in Sec. 5.1, in the discussion of electron correlation. If we had chosen to follow the original discussion of Gell-Mann and Brueckner (1957), the correlation energy could be derived without first discussing the electron self-energies. The point is made again that the collective properties are not significantly dependent on first deriving the one-electron properties. The discussion of electron correlation followed Quinn and Ferrell (1958) and first considered the electron self-energy. The results of that discussion can be used as the starting point for the present derivations. The best calculations which have been reported of single-electron properties calculate only the screened exchange interaction.

The electron self-energy from the screened exchange energy is

$$\Sigma_{sx}(\mathbf{k}, ik_n) = -\frac{1}{\beta}\sum_{iq_m}\int\frac{d^3q}{(2\pi)^3}v_q\frac{\Gamma(\mathbf{q}, iq_m)}{\varepsilon(\mathbf{q}, iq_m)}\mathscr{G}^{(0)}(\mathbf{k}+\mathbf{q}, ik_n + iq_m) \tag{5.317}$$

$$\varepsilon(\mathbf{q}, iq_m) = 1 - v_q P^{(1)}(\mathbf{q}, iq_m)\Gamma(\mathbf{q}, iq_m) \tag{5.318}$$

$$\Gamma(\mathbf{q}, iq_m) = \frac{1}{1 + v_q G_H(q) P^{(1)}(\mathbf{q}, iq_m)} \tag{5.319}$$

The integrand contains a vertex function $\Gamma(\mathbf{q}, iq_m)$, as discussed in Sec. 5.5. Calculations using (5.317) are shown in Fig. 5.21. Results are obtained by a numerical evaluation of the multidimensional integral. Figure 5.21(a) shows the real part of the self-energy Σ_{sx} as a function of k for r_s values appropriate to common metals. All of the curves have a similar shape. The strong dip in the curve is due to the coupling of the electron to the plasmon. The imaginary self-energy $\mathrm{Im}[\Sigma_{sx}]$ has a big jump in magnitude when the electron has enough kinetic energy to emit a plasmon. The Kramers–Kronig relation between $\mathrm{Re}[\Sigma]$ and $\mathrm{Im}[\Sigma]$ guarantees that $\mathrm{Re}[\Sigma]$ has a dip when $\mathrm{Im}[\Sigma]$ has a jump.

A similar dip in the self-energy is obtained whenever the electron couples strongly to an oscillator. The self-energy of electrons coupled to LO phonons in insulators has the same shape, as discussed in Problem 7.7. Of course, there the dip happens at the energy characteristic of the phonon, while here it occurs at the energy of the plasmon.

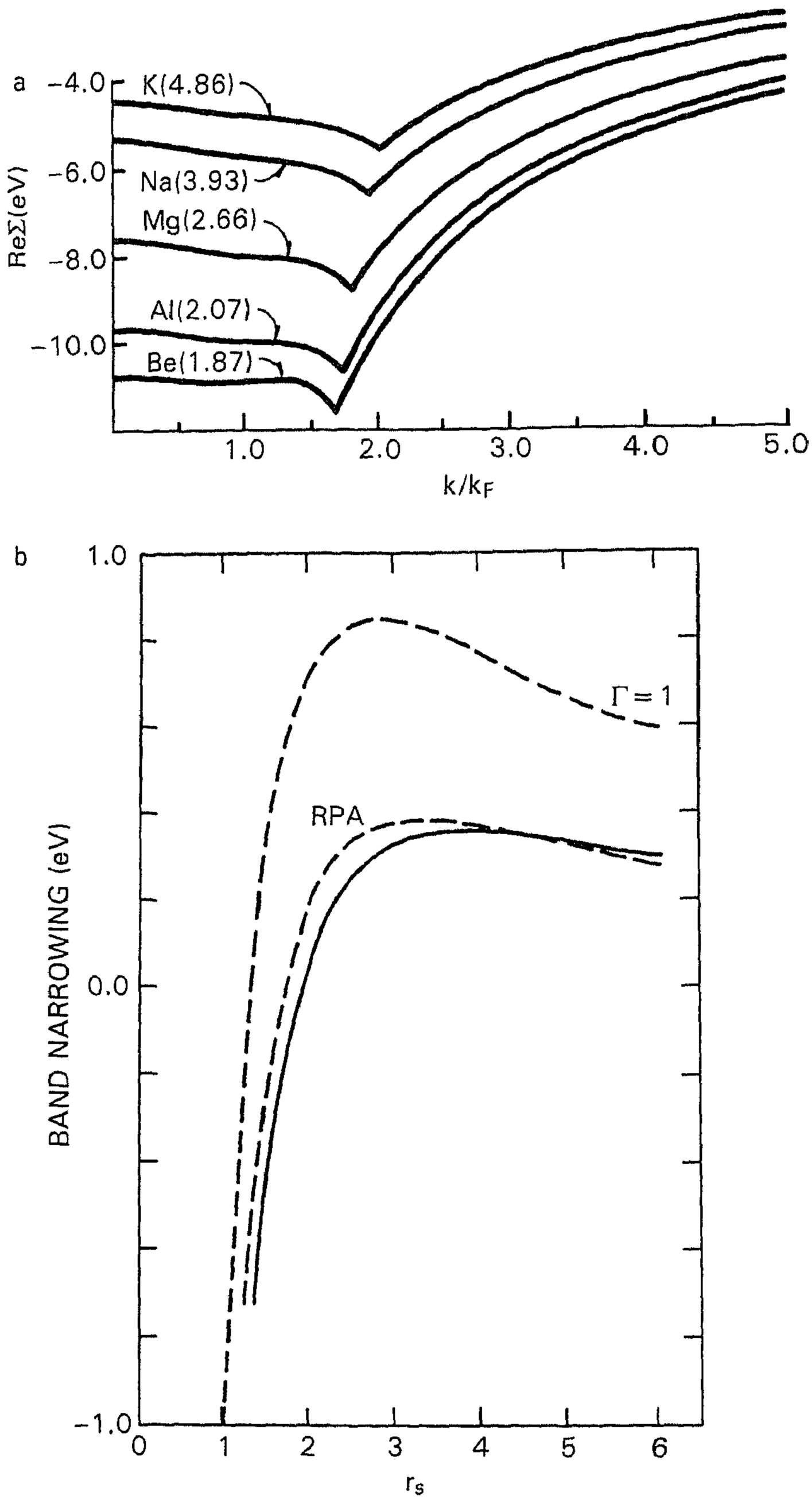


FIGURE 5.21 (a) The real part of the screened exchange energy $\Sigma_{sx}(k, \xi_k)$ as a function of k for values of r_s appropriate for simple metals. The dip is caused by plasmon emission. *Source*: Shung *et al.* (1987). (b) The change in energy bandwidth for electrons interacting with electron–electron interactions. The RPA result ($G = 0$) is compared to the $G_H(q)$ of Vashishta and Singwi. The results are quite similar. *Source*: Mahan and Sernelius (1989).

The electron self-energy also contributes to the energy width of the occupied band of electrons. This change in width can be crudely estimated as the difference between the self-energy calculated at $k = 0$ and at $k = k_F$.

$$\Delta W = \Sigma_{sx}(k_F, 0) - \text{Re}\, \Sigma_{sx}(0, -E_F) \tag{5.320}$$

The energy $\xi = 0$ at the Fermi energy, and equals $-E_F$ at the bottom of the band. Figure 5.21(b) compares ΔW calculated using RPA and the Vashishta–Singwi (VS) form for $G_H(q)$

given in Table 5.2. The results are quite similar. The local field factor of $G_H(q)$ has only a small influence upon the bandwidth of the electrons. The curve labeled $\Gamma = 1$ uses $G_H(q)$ in $\varepsilon(q, \omega)$ but not in Γ. This inconsistent theory is called the "GW" approximation.

5.8.1. Renormalization Constant Z_F

The renormalization coefficient of the one-electron Green's function is defined for a point $(k, \varepsilon = \xi_k)$ as [see (3.155)]

$$Z(k) = \frac{1}{[1 - \partial\Sigma_{sx}(k, \varepsilon)/\partial\varepsilon]_{\varepsilon=\xi_k}} \tag{5.321}$$

Usually the retarded self-energy Σ_{sx} is complex, as is the renormalization coefficient. At the particular point, $k = k_F$ and $\xi = 0$, the self-energy Σ_{sx} is exactly real, since $\mathrm{Im}(\Sigma)$ vanishes. Electrons on the Fermi energy have an infinitely long mean free path, or lifetime. They have no unoccupied electron states into which they might scatter. The renormalization coefficient is real at this point. At the Fermi energy it is called $Z_F = Z(k_F)$

$$Z_F = \frac{1}{[1 - \partial\Sigma_{sx}(k_F, \varepsilon)/\partial\varepsilon]_{\varepsilon=0}} \tag{5.322}$$

The renormalization coefficient is interpreted as the amount of single-particle behavior of the particle-like excitation in the electron gas. Since the excitations behave like particles, they are called *quasiparticles*. Generally the spectral function $A(k, \varepsilon)$ has a central peak plus some smooth background wings. This behavior is illustrated schematically in Fig. 5.22. The renormalization coefficient is defined as the area under the central peak of the spectral function. If the peak is broad, then the definition is ambiguous, since the limits of the central peak are ill-defined. Precisely at the Fermi surface the quasiparticle central peak is a sharp delta function, because $\mathrm{Im}\,\Sigma = 0$. Here the renormalization coefficient has a precise meaning: the spectral weight under the delta function peak. As discussed in Sec. 3.3, the renormalization coefficient Z_F must be less than or equal to unity since the total area under $A(k, \varepsilon)$ is unity.

The quantity Z_F has an additional interpretation, which can be measured experimentally. The momentum distribution of the electron gas is given by [see (3.135)]

$$n_{\mathbf{k}} = \int \frac{d\varepsilon}{2\pi} A(\mathbf{k}, \varepsilon) n_F(\varepsilon) \tag{5.323}$$

Actual results for an interacting electron gas have the form illustrated by the solid line in Fig. 5.23 (Daniel and Vosko, 1960). The dashed line is for a noninteracting Fermi system, which has all the particles below $k = k_F$ at $T = 0$. For the interacting Fermi system, even at $T = 0$,

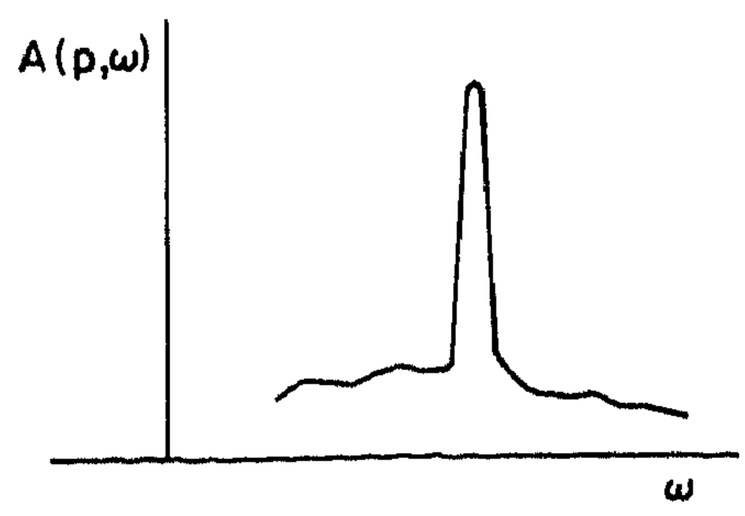


FIGURE 5.22 Spectral function of A.

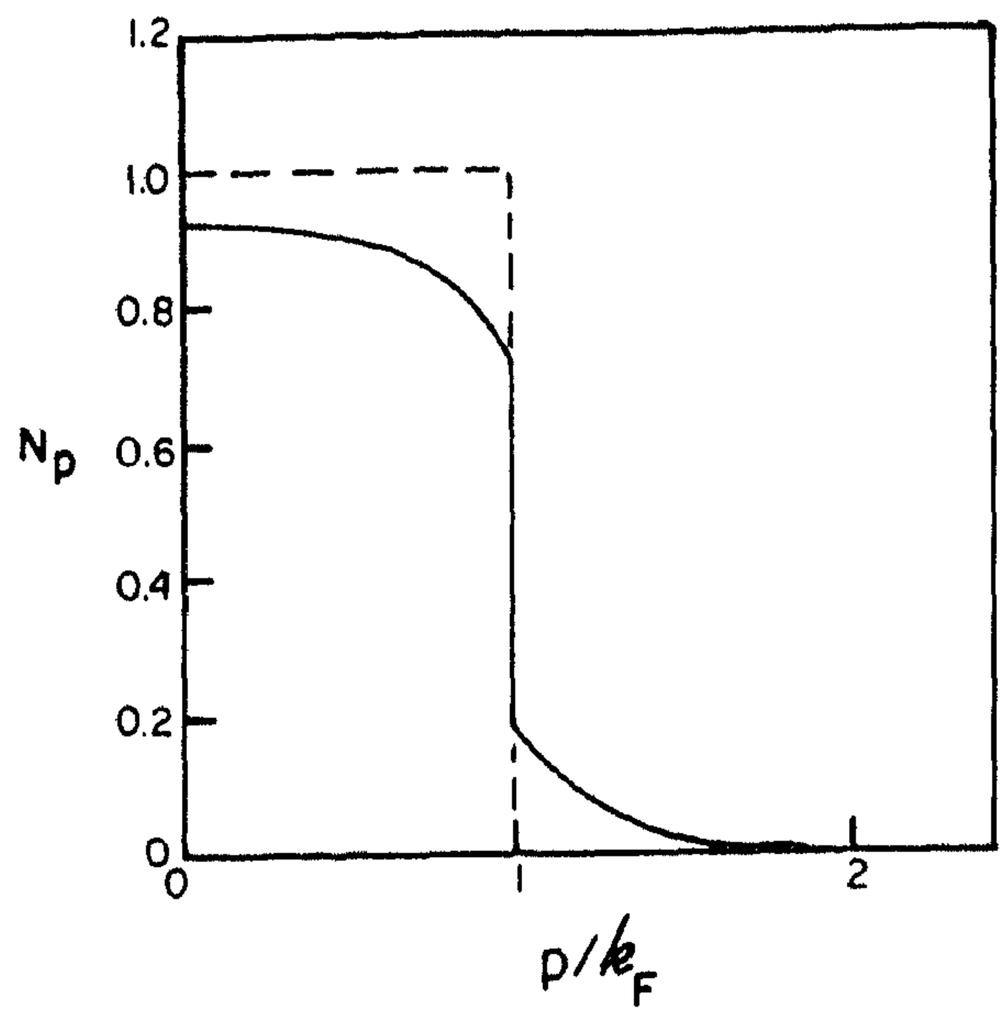


FIGURE 5.23 Momentum density of the homogeneous electron gas for $r_s = 3.97$. *Source*: Daniel and Vosko (1960) (used with permission).

the momentum distribution is the solid line, with some components for $k > k_F$. The quantity n_k should be distinguished from the energy distribution $n_F(\varepsilon) = 1/(e^{\beta\varepsilon} + 1)$, which is a sharp step at zero temperature. The renormalization coefficient Z_F is the magnitude of the step at $k = k_F$ in the momentum distribution n_k. This assertion is easily shown using the definition of n_k and the fact that

$$A(k_F, \varepsilon) = 2\pi Z_F \delta(\varepsilon) + f(\varepsilon) \tag{5.324}$$

where $f(\varepsilon)$ is a smooth function. The step in the momentum distribution in the Fermi gas may be measured in actual metals by Compton scattering.

There have been two major types of calculations with these formulas. The first is the RPA, where we use Hedin's (1965) results. The other is the revised Hubbard model work of Rice (1965). The two calculations of Z_F are shown in Table 5.4 for different values of r_s. Rice only reported values of $r_s < 4$. The two calculations are in very good agreement. The two models of the dielectric function make the same prediction regarding this quantity.

Eisenberger *et al.* (1972) reported Compton scattering experiments in metallic sodium and lithium which demonstrated the existence of the high momentum tail in n_k. Compton scattering is the inelastic scattering of photons by the electrons in the solid. The photons are scattered by all the electrons, both in core states and the conduction band. The conduction band profiles, which are our present interest, are obtained after subtracting the core scattering.

TABLE 5.4 Z_F as calculated by (*a*) Hedin (1965) and (*b*) Rice (1965)

r_s	RPA[a]	Hubbard[b]
0	1	1
1	0.859	0.87
2	0.768	0.77
3	0.700	0.70
4	0.646	0.63
5	0.602	
6	0.568	

The best results are obtained for metals with low atomic number, since they have fewer core levels to subtract. Eisenberger *et al.* analyzed their data in terms of the *sudden approximation.* This theory assumes that the scattering rate of photons is proportional to

$$\frac{d^2\sigma}{d\omega d\Omega} \propto \int \frac{d^3p}{(2\pi)^3} n_p \delta(\omega + \varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p+q}}) \tag{5.325}$$

where ω is the energy transferred by the photon and $\mathbf{q}$ is the momentum transferred. The electrons are scattered from $\mathbf{p}$ to $\mathbf{p} + \mathbf{q}$, and the average is taken over the initial distributions n_p of the interacting electron system. When the angular integral is done to eliminate the delta function, it is found that the experimental quantity is only a function of the combination of variables called $Q = (m/q)(\varepsilon_q - \omega)$:

$$\int_{-1}^{1} d\nu \delta\left(\omega - \varepsilon_q - \frac{pq\nu}{m}\right) = \frac{m}{pq}\Theta(p - |Q|)$$

$$\frac{d^2\sigma}{d\omega d\Omega} \propto J(Q) \tag{5.326}$$

$$J(Q) = \int_{|Q|}^{\infty} p dp n_p \tag{5.327}$$

For a noninteracting electron gas, one has that $n_p = \Theta(k_F - p)$ and the scattering function $J(Q)$ is an inverted parabola:

$$J(Q) = \tfrac{1}{2}(k_F^2 - Q^2)\Theta(k_F - |Q|) \tag{5.328}$$

Thus $J(Q)$ is shown as the dashed line in Fig. 5.24(a). The solid line shows the type of behavior expected for an interacting electron gas. There is a long tail at high values of Q, which arises from the similar high tail in n_p for values of $p > k_F$.

The experimental and theoretical results of Pandey and Lam (1973) are shown in Fig. 5.24(b). The triangular points are the experiment, and they are joined by the solid line. The

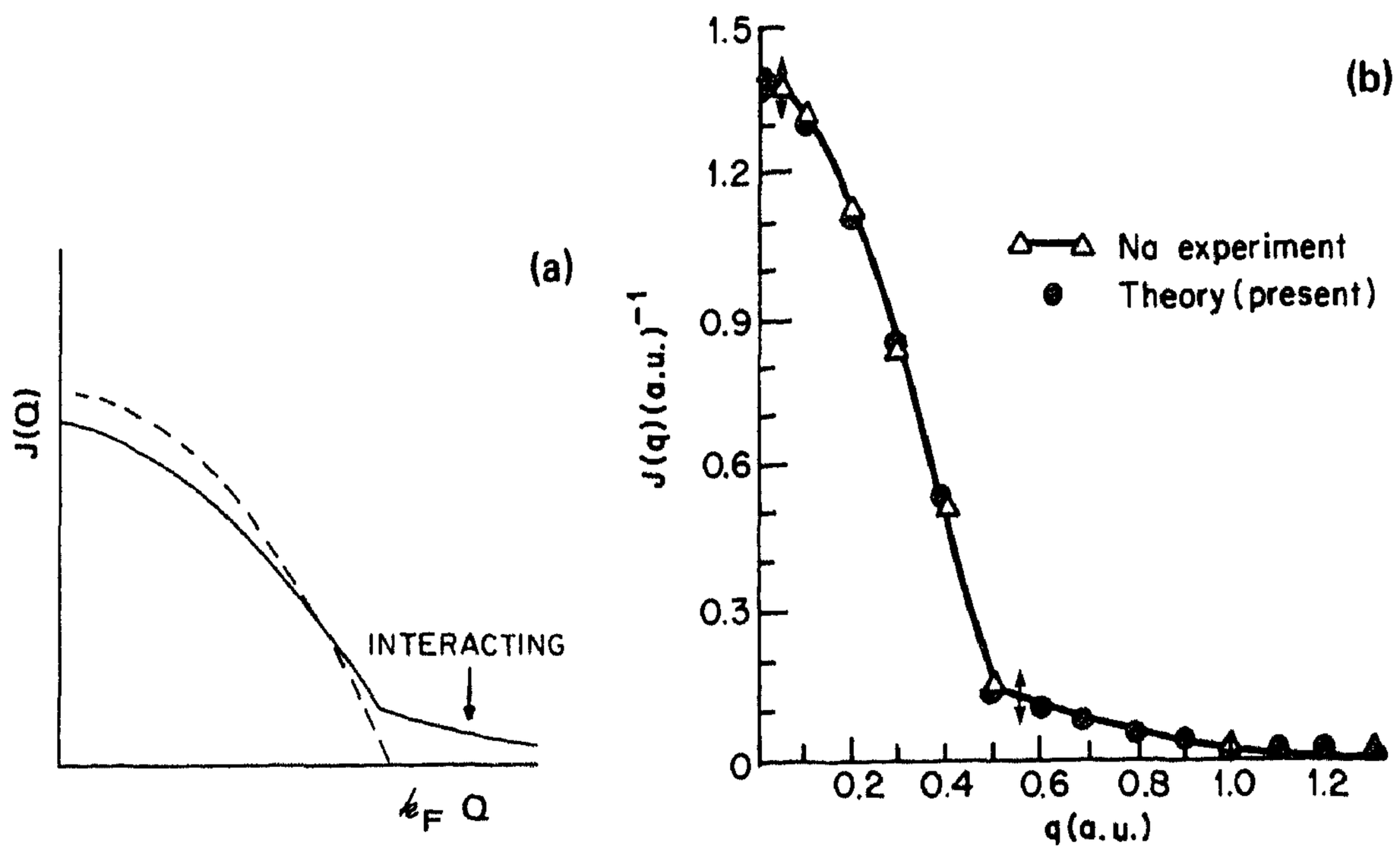


FIGURE 5.24 (a) The dashed line shows the momentum distribution in Compton scattering from a noninteracting electron gas, and the solid line is the prediction for an interacting system. (b) The experimental results in metallic sodium (solid points) compared with the theory (open triangles) for this metal. *Source*: Pandey and Lam (1973) (with permission).

circular points are the theory. The agreement is obviously excellent. Sodium is the ideal metal for testing the distribution n_k since its Fermi surface is quite spherical. Other experimentalists (Eisenberger *et al.*, 1972) find that the Compton scattering depends only upon Q, which can be tested by different choices of q and ω.

5.8.2. Effective Mass

The formula for the effective mass was given previously in (3.160). It will be evaluated for electrons on the Fermi surface of the metal:

$$\left(\frac{m}{m^*}\right) = \left|\frac{1 + \partial\Sigma(k, \varepsilon)/\partial\varepsilon_k}{1 - \partial\Sigma(k, \varepsilon)/\partial\varepsilon}\right|_{k=k_F, \varepsilon=0} = Z_F\left[1 + \frac{\partial\Sigma(k, \varepsilon)}{\partial\varepsilon_k}\right]_{k=k_F, \varepsilon=0} \tag{5.329}$$

The results of Rice were obtained by a slightly different formula, which should give similar results. His values are shown in Table 5.5. The ratio m/m^* is less than unity for values of r_s less than 2, while it is slightly larger than unity for larger values of r_s. The effective masses of electrons on the Fermi surface of the metal may be measured quite accurately by many types of cyclotron resonance experiments. The effective mass values for real metals are determined by three major contributions. The first is the electron–electron interactions, and they make only a small contribution to the mass. The second is due to band structure, which is quite variable. It is small in sodium but large in lithium and many other metals. The third contribution is due to electron–phonon interactions, which is quite sizable. A comparison between theory and experiment is deferred until the phonon contribution is calculated in Sec. 7.4.

Another application of the effective mass theory is in the low-temperature specific heat, or heat capacity. The free-electron theory of metals predicts that the specific heat of a metal, at low temperature, is linear in temperature. A free-electron gas of N_e particles has the result

$$C_{v0} = \frac{3\pi^2}{4} N_e k_B^2 T \frac{m}{(3\pi^2 n_0)^{2/3}\hbar^2} \tag{5.330}$$

For an interacting system, the prediction is that the free-electron mass is replaced by the effective mass at the Fermi energy m^*. The ratio of the linear term in the specific heat, at low temperature, is just the ratio of the effective masses:

$$\frac{C_v}{C_{v0}} = \frac{m^*}{m} \tag{5.331}$$

The quantity C_{v0}/C_v is calculated in a similar manner as m/m^*. This ratio is also influenced by phonon effects.

TABLE 5.5 Results of Rice (1965) for m/m^* using the Hubbard dielectric function

r_s	Hubbard
0	1
1	0.96
2	0.99
3	1.02
4	1.06

5.8.3. Mean-Free-Path

The mean-free-path l_k is the average distance the electron travels between scattering events. Similarly, the lifetime τ_k is the average time between scattering events and is related by $l_k = v_k \tau_k$. In the electron gas, most scattering processes are inelastic, and the electron gains or loses energy each time it scatters. These quantities are defined in terms of the imaginary part of the self-energy:

$$\frac{\hbar}{\tau_k} = -2\ \mathrm{Im}[\Sigma(k, \xi_k)] \tag{5.332}$$

$$\frac{1}{l_k} = -\frac{2}{\hbar v_k}\mathrm{Im}[\Sigma(k, \xi_k)] \tag{5.333}$$

Most calculations have been done with the screened exchange energy in (5.56). The rate of scattering is proportional to $S(q, \omega) = \mathrm{Im}[1/\varepsilon(q, \omega)]$. The scattering has an energy transfer $\omega = \xi_{\mathbf{k}+\mathbf{q}} - \xi_{\mathbf{k}}$, where the initial state has $\xi_{\mathbf{k}} > 0$ and the final state $\xi_{\mathbf{k}+\mathbf{q}} > 0$ is empty. A simple expression for the mean-free-path is

$$\frac{1}{l_k} = -\frac{2}{\hbar v_k}\int \frac{d^3q}{(2\pi)^3} v_q[1 - n_F(\xi_{\mathbf{k}+\mathbf{q}})]\mathrm{Im}\left[\frac{\Gamma_H(q, \xi_{\mathbf{k}+\mathbf{q}} - \xi_{\mathbf{k}})}{\varepsilon(q, \xi_{\mathbf{k}+\mathbf{q}} - \xi_{\mathbf{k}})}\right] \tag{5.334}$$

This formula has formed the basis for most treatments of electron scattering in metals. The expression is expected once $\mathrm{Im}[1/\varepsilon(q, \omega)]$ is identified as the rate of making excitations $(\mathbf{q}, \omega)$ in the solid.

For small values of initial energy ξ_k it is necessary to numerically integrate this expression. It is a double integral, since one must do both the dq and $d\nu$ ($\nu = \cos\theta$) integrations. This calculation has been done by Lundqvist (1969). For more energetic particles, the excitation spectrum of $\mathrm{Im}[1/\varepsilon(q, \omega)]$ is dominated by long-wavelength plasmons. One way to evaluate this expression approximately is to assume that the plasmons totally dominate the excitation spectrum. For small q use the limiting form given in (5.311)–(5.314).

An analytical result may be derived for the mean-free-path. Insert the delta function for $\mathrm{Im}(1/\varepsilon)$ into the integrand and change integration variables to $\mathbf{p} = \mathbf{k} + \mathbf{q}$:

$$\frac{1}{l_k} = \frac{e^2\omega_p}{\hbar v_k}\Theta(\xi_k - \omega_p)\int_{k_F}^{\infty} p^2 dp\,\delta(\varepsilon_k - \omega_p - \varepsilon_p)\int_{-1}^{1} d\nu \frac{1}{k^2 + p^2 - 2pk\nu} \tag{5.335}$$

The factor $\Theta(\xi_k - \omega_p)$ arises from energy conservation $\xi_k - \omega = \xi_{\mathbf{k}+\mathbf{q}} > 0$. The integrals may be evaluated:

$$\frac{1}{l_k} = \frac{1}{2a_0}\left(\frac{\omega_p}{\varepsilon_k}\right)\Theta(\xi_k - \omega_p)\ln\left[\frac{\sqrt{\varepsilon_k} + \sqrt{\varepsilon_k - \omega_p}}{\sqrt{\varepsilon k} - \sqrt{\varepsilon_k - \omega_p}}\right] \tag{5.336}$$

The formula is given in terms of the kinetic energy ε_k measured from the bottom of the conduction band, although plasmons are emitted only when $\xi_k > \omega_p$ or $\varepsilon_k > \omega_p + E_F$. Equation (5.336) is almost identical to an old formula of Bethe (1930) for the rate of energy loss of charged particles in solids.

Figure 5.25 shows a comparison of these theories with the experiments in aluminum. The metal aluminum is chosen as a basis of comparison because the best experimental results are available of any free-electron metal. The data is due to Tracy (1974). The plasmon formula (5.336) only needs the aluminum plasmon energy of $\hbar\omega_p = 15.5$ eV to determine the dashed line. The plasmon theory in Fig. 5.25 is too low by about 30–40%. The electron mean-free-path is longer than would be predicted by the simple plasmon formula. The

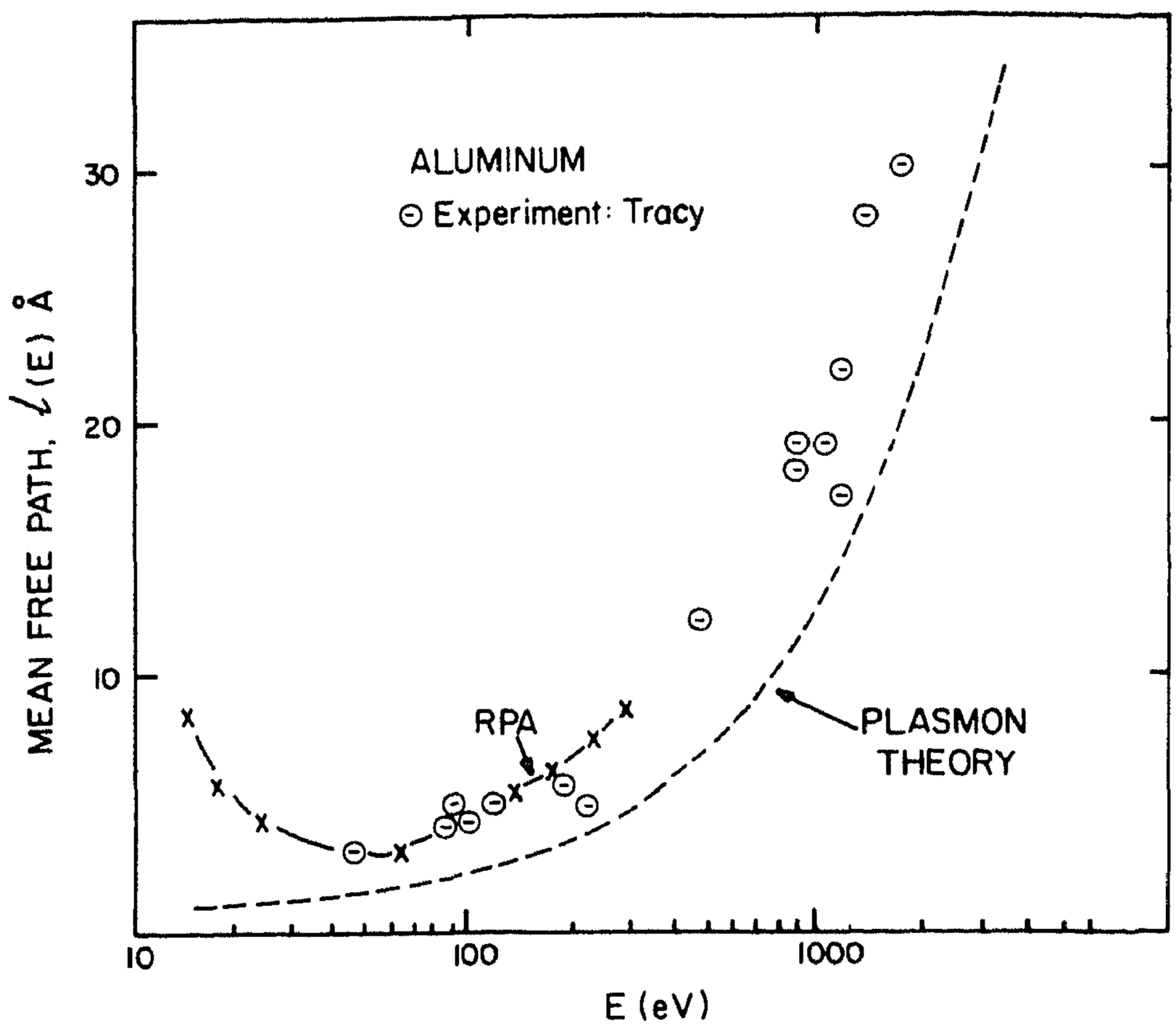


FIGURE 5.25 The mean-free-path of electrons in aluminum as a function of energy above the Fermi surface. Circles are data points from Tracy (1974). The dashed line with × points is the RPA calculation of Lundqvist (1969). The other dashed line is the plasmon theory which ignores pair production, and its predictions are inaccurate.

plasmon theory is unreliable at lower energies. Experiments show that the rate of making plasmons vanishes below 40 eV (Flodstrom *et al.*, 1977). The other theoretical curve is the RPA result, which used the calculations of Lundqvist (1969) for $r_s = 2$, which is close to the aluminum value of $r_s = 2.07$. He numerically evaluated (5.334). His curve is in excellent agreement with the experimental results. Similar measurements have been done on other metals, although few on free-electron metals. See the reviews by Echenique *et al.* (2000).

Another experiment is to measure the inelastic scattering of electrons by thin metal foils. The electrons are shot through the foil, and their cross section for inelastic scattering is measured as a function of angle and energy. The electrons must be fairly energetic to penetrate the foil. The experiments show that the electrons mostly lose energy by exciting long-wavelength plasmons (Marton *et al.*, 1962).

First calculate the angular dependence of the scattering when the electron emits one long-wavelength plasmon. The process is that the energetic electron, usually in the kilovolt range, loses one plasmon of energy and is scattered through an angle θ. The energetic electron scatters from **k** to **p**. The scattering rate may be obtained from (5.335) by eliminating the angular integrals. The result is the differential scattering rate as a function of angle:

$$\frac{dw}{d\Omega} = \frac{e^2\omega_p}{2\pi}\int_{k_F}^{\infty} p^2 dp \frac{\delta(\varepsilon_k - \omega - \varepsilon_p)}{k^2 + p^2 - 2pk\cos(\theta)} \tag{5.337}$$

The delta function determines the value of p. For energetic electrons, this momentum is approximately

$$p = \sqrt{k^2 - 2m\omega_p} \approx k - \frac{\omega_p}{v_k} \tag{5.338}$$

The typical scattering angle is very small, and it is possible to use the small-angle approximation for $\cos\theta = 1 - \theta^2/2$. The denominator can be simplified to

$$k^2 + p^2 - 2pk\cos(\theta) \approx (k-p)^2 + kp\theta^2 \approx k^2(\theta_0^2 + \theta^2) \tag{5.339}$$

$$\theta_0^2 = \frac{\omega_p}{2\varepsilon_k} \tag{5.340}$$

and the cross section has the simple form derived by Ritchie (1957):

$$\frac{dw}{d\Omega} = \frac{e^2\omega_p}{2\pi v_k}\frac{1}{\theta_0^2 + \theta^2} \tag{5.341}$$

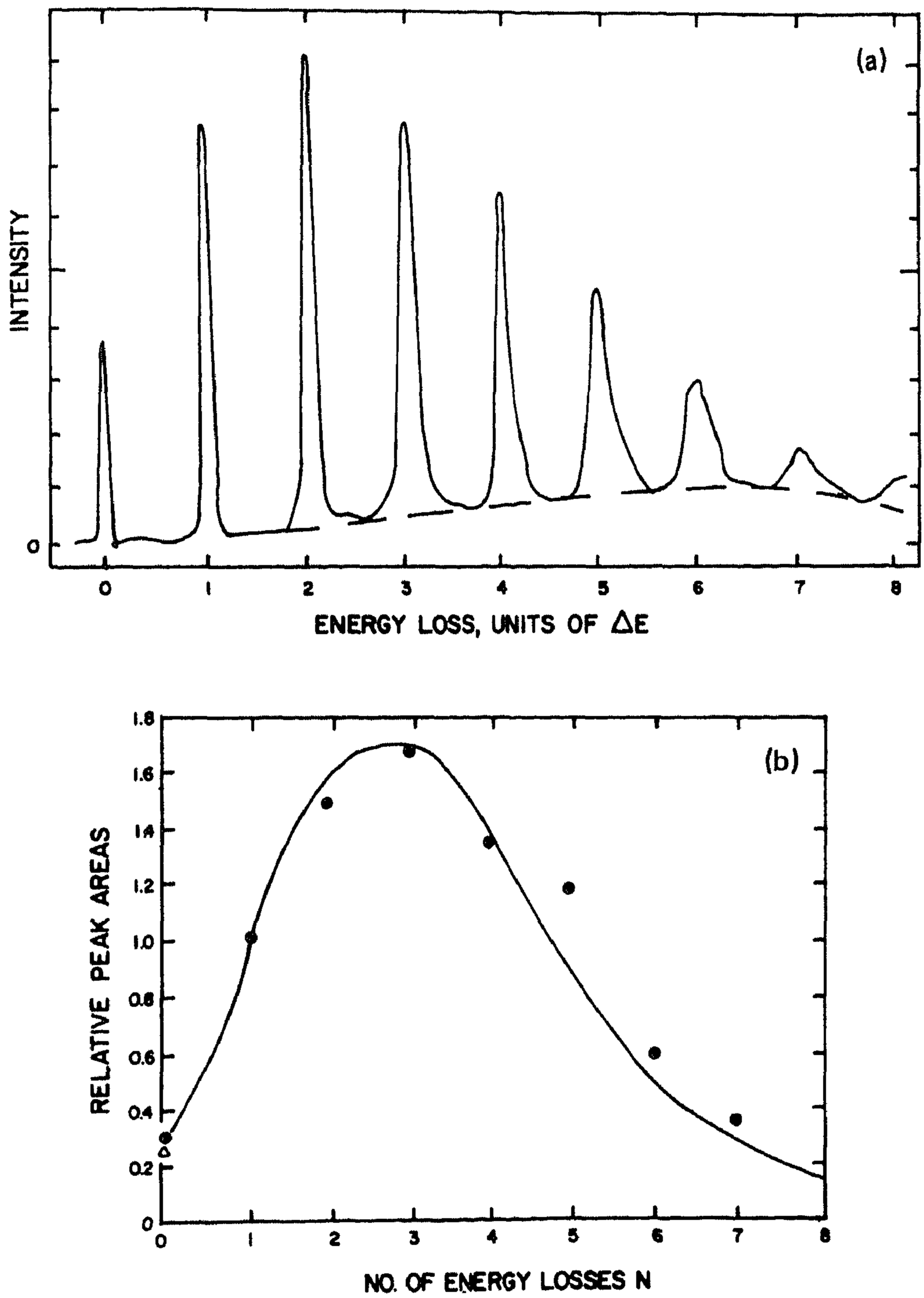


FIGURE 5.26 Energy loss spectrum at $\theta = 0$ for electrons going through aluminum films. The relative peak heights are shown in (b) and are well fit by a Poisson distribution shown by the solid line. *Source*: Marton *et al.* (1962) (used with permission).

The scattering distribution is Lorentzian. The angular width θ_0 is typically a fraction of a degree when ε_k is a kilovolt. The angular distribution is sharply peaked in the forward direction. This behavior is found experimentally (Sueoka, 1965).

Another aspect of electron energy loss through metal foils is shown in Fig. 5.26, which is the data of Marton *et al.* (1962) for aluminum. The peak on the left of Fig. 5.26(a) is from the electrons which go through the foil with no energy loss. The next peak is from those which excite one plasmon, and the nth is from electrons which excite $n-1$ plasmons. The angular dependence of these peaks was measured and found to be Lorentzian. Another feature of these peaks is shown in Fig. 5.26(b), which is the peak intensity, or the area under each curve. The background is subtracted according to the dashed line in Fig. 5.26(a). The peak intensities follow a Poisson distribution (solid line), so that the intensity of the nth peak is given by

$$P_n = e^{-\alpha}\frac{\alpha^n}{n!}, \qquad \alpha = \frac{d}{l} \tag{5.342}$$

The parameter α is the thickness of the film d divided by the mean-free-path for plasmon emission. The Poisson distribution is a characteristic of the independent boson model of Sec. 4.3. A Poisson distribution occurs whenever the bosons, in this case plasmons, are emitted independently of previous emissions. This model is reasonable for these fast electrons. The electron energy was 20 keV. An electron of this energy does not lose a significant fraction of its energy when it emits a plasmon of 15 eV. Similarly, its angle does not change appreciably, since the scattering is mostly in the forward direction. The electron trajectory is largely unchanged by the plasmon emissions. All emission events have the same likelihood and are independent of previous emissions. It makes sense that the independent boson model should be applicable here. Of course the zero-temperature version of the theory is used, since even at room temperatures, the number of plasmons thermally present $1/(e^{\beta\omega_p}-1)$ is negligible.

Problems

1. Consider the ground state of the ferromagnetic electron gas. The particles are plane waves, but all spins point in the same direction.

   (a) Find the ground state energy of the system in the Hartree–Fock approximation.
   (b) Compare this with the paramagnetic state (equal numbers of up and down spins) calculated in the same approximation. Are there values of r_s for which the ferromagnetic state is lower in energy?

2. In the Wigner lattice the electron feels a harmonic potential $V(r) = -3/r_s + r^2/r_s^3$. Find the zero-point energy from oscillations in this well by assuming that each electron moves independently of the others.

3. The correlation energy may also be evaluated by Rayleigh–Schrödinger perturbation theory (Gell-Mann and Brueckner, 1957). The second-order ground state energy is

$$E_g^{(2)} = \sum_I \frac{|\langle I|V|i\rangle|^2}{E_i - E_I} \tag{5.343}$$

Obtain the ground state energy terms which contribute in second order, and write them down in terms of wave vector summations.

4. Calculate the equilibrium value of r_{s0} for metallic hydrogen using NP correlation. Also calculate the ground state energy per electron.

5. Calculate the cohesive energy of metallic helium assuming it is a uniform electron gas, with α particles spaced on a crystal lattice. Compare your result to atomic helium.

6. Consider putting a single point charge Q into a homogeneous electron gas. Show that to order Q^2 the energy required to do this is

$$\Delta\Omega = \frac{1}{2}\int \frac{d^3q}{(2\pi)^3}\frac{4\pi Q^2}{q^2}\left[\frac{1}{\varepsilon(q)} - 1\right] + O(Q^3) \tag{5.344}$$

Estimate this result for the Thomas–Fermi model.

7. Use (5.117) to estimate $S(q)$ at large q in the RPA. At large q, one has that $\mathrm{Im}(1/\varepsilon) \approx -\varepsilon_2$.

8. Use (5.117) to find the behavior of $S(q)$ as $q \to 0$.

9. Set $1 - g(r) = A\exp(-rk_F)$.

   (a) Determine A by the normalization condition (5.71).
   (b) Determine $S(q)$.
   (c) Determine $G(q)$. Compare this on a piece of graph paper with Hubbard and Singwi–Sölander for $r_s = 3$.

10. Calculate the compressibility of the electron gas from the ground state energy:

   (a) in the Hartree–Fock approximation.
   (b) including both HF and Wigner correlation energy.

Plot both results on a piece of graph paper as K_f/K vs. r_s for metallic densities.

11. Start from the Singwi–Sjölander expression (5.198) for $G_H(q)$. Take the limit that $q \to \infty$ and find their prediction for this limit.

12. Calculate $S(q)$ in the Hartree–Fock approximation. First show that the starting point is $S(q) = 1 - (2/N)\sum_{\mathbf{k}} n_{\mathbf{k}} n_{\mathbf{k}+\mathbf{q}}$.

13. Calculate $\gamma(r_s)$ in the Hartree–Fock approximation using the results of Problem 12. Verify that the correlation energy (5.249) is zero with this choice of $\gamma(r_s)$.

14. Use the form of the dielectric function (5.187) to show that the plasma dispersion relation is

$$\omega_p(q) = \omega_p(0) + \alpha\left(\frac{\hbar q^2}{m}\right) + O(q^4) \tag{5.345}$$

$$\alpha = \frac{3}{10}\frac{\hbar k_F^2}{m\omega_p}\left[1 - \frac{5}{9}a\left(\frac{q_{TF}}{k_F}\right)^2\right] \tag{5.346}$$

where a is the parameter which enters into the compressibility (5.270).

15. Prove the ground state energy theorem (5.242) starting from (3.297) using just the electron–electron interaction.

16. In the Vashishta–Singwi theory, the constant a in the compressibility is no longer equal to γ as given by (5.244). Find the new relationship to γ.

17. Does the RPA satisfy the sum rule (5.307)? Show by explicitly doing the integrations.

18. Calculate $\varepsilon_2(\mathbf{q}, \omega)$ in RPA for a classical gas which obeys Maxwell–Boltzmann statistics.

19. Derive the sum rule for the third moment for a Hamiltonian of particles in a potential field $V(\mathbf{r})$. Do this by evaluating the expectation of the triple commutator with H:

$$H = \sum_{\mathbf{p}\sigma} \xi\mathbf{p} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \sum_{\mathbf{q}} \rho(\mathbf{q}) V(\mathbf{q}) \tag{5.347}$$

$$\langle \omega^3 \rangle = \langle [[H, [H, [H, \rho(\mathbf{q})]]], \rho(-\mathbf{q})] \rangle \tag{5.348}$$

20. Prove that the step in n_k at $k = k_F$ is Z_F.

21. The screened exchange energy (5.317) may be calculated exactly in the Thomas–Fermi approximation. Do this, and use your answer to obtain analytical expressions for

(a) the self-energy at $p = 0$.
(b) the self-energy at $p = k_F$.
(c) the effective mass at the Fermi energy.

Plot all three results as a function of r_s on a piece of graph paper.

22. Calculate $\varepsilon(\mathbf{q}, \omega)$ in RPA for the two-dimensional electron gas at zero temperature. Show that the plasmon dispersion can be given by an analytical expression.

23. Calculate the average kinetic and exchange energies per electron for a two-dimensional electron gas with Coulomb interactions e^2/r. Define r_s by $\pi n_0 a_0^2 r_s^2 = 1$, and your answer should be $E_g = 1/r_s^2 - 1.20/r_s \ldots$. Find the analytical form of the coefficient of the exchange energy (it is not 6/5). For a review of correlation calculations, see Jonson (1976).

24. In two dimensions show that the exchange energy at the Fermi surface is $\Sigma_x = -2e^2 k_F/\pi$. Use this result to calculate the compressibility in the Hartree–Fock approximation. At what density r_s does the compressibility go negative in two dimensions?

25. Draw the connected Feynman diagrams for the electron self-energy from electron–electron interactions that contain two or three internal Coulomb lines. Show that all terms can be written as $v_q \Gamma/\varepsilon$ in (5.317). Which diagrams contribute to Γ and which to ε?

Chapter 6

Strong Correlations

The previous chapter discussed metals in which correlation plays a minor role in the dynamics of the electron. The properties of the electron gas could be calculated simply and accurately. The present chapter discusses other metallic systems in which correlation plays a much larger role in the dynamics of the electron. These topics come under the title of *strong correlation*. Strong correlation causes itinerant magnetism.

The theory of magnetism in metals continues to be one of the challenging subjects of modern theoretical physics. Magnetic phenomena are important, and are observed in a wide variety of materials. The equations are difficult to solve. A variety of theoretical techniques are applied to magnetic phenomena, such as Monte Carlo, renormalization group, and Green's functions. Only the latter technique is discussed here.

Strong correlation is usually treated by solving the Anderson model or the Hubbard model. They are discussed here, along with the Kondo model (1964) which was the historical precedent. Most Green's function analyses have focused on these three model Hamiltonians. Each of these three models is not a single mathematical problem. Instead, each has a number of variations. For example, the original Kondo model describes a system of conduction electrons interacting with a single localized spin. The local spin can have any value of angular momentum S, and the conduction electrons can be treated in one, two, or three dimensions. Recent variations include the dense or periodic Kondo problem, where there are N local spins arranged on a lattice. Another feature of all three models is that they can be solved exactly in one dimension, using the Bethe Ansatz, but not in higher dimension. One-dimensional results are not a useful guide to collective effects in higher dimension, since there are neither phase transitions nor long-range order in one dimension at nonzero temperatures.

6.1. KONDO MODEL

For many years experimentalists noticed that magnetic impurities in nonmagnetic metals caused anomalous behavior in the low-temperature resistivity $\rho(T)$. Magnetic impurities are those with a moment caused by partially filled d- or f-electron shells. An example is manganese impurities in copper. The electron scattering from a nonmagnetic impurity makes a contribution to the resistivity that is independent of temperature. However, a magnetic impurity causes a resistance minimum at a nonzero temperature. Kondo explained this behavior as due to spin-flip scattering between the conduction electrons and the localized

spin. The resistance minimum is now called the *Kondo effect*. The *Kondo model* is described by the Hamiltonian

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} C^{\dagger}_{\mathbf{k}\sigma} C_{\mathbf{k}\sigma} - \frac{J}{N} \sum_{j\mathbf{kp}} e^{i\mathbf{R}_j \cdot (\mathbf{k}-\mathbf{p})} [(C^{\dagger}_{\mathbf{k}\uparrow} C_{\mathbf{p}\uparrow} - C^{\dagger}_{\mathbf{k}\downarrow} C_{\mathbf{p}\downarrow}) S^{(z)}_j + C^{\dagger}_{\mathbf{k}\uparrow} C_{\mathbf{p}\downarrow} S^{(-)}_j + C^{\dagger}_{\mathbf{k}\downarrow} C_{\mathbf{p}\uparrow} S^{(+)}_j] \tag{6.1}$$

The first term represents the kinetic energy of the conduction electrons. If the local spin is from a d-shell, the conduction band is usually formed from atomic orbitals that have s- and p-symmetry. If the local spin is from an f-orbital, then the conduction band could be from s, p, or d electrons.

The remaining terms represent the scattering from the local spin at the site $\mathbf{R}_j$. The number of local spins is N_i, and the concentration is $c = N_i/N$, where N is the number of atoms in the solid. Usually it is assumed that c is quite small. The electron self-energy is calculated to order $O(c)$. This approximation is equivalent to assuming that the conduction electron scatters from one impurity at a time, or that the impurities are widely separated.

The interaction term can be written compactly as $\mathbf{s} \cdot \mathbf{S}$, where the small $\mathbf{s}$ is the spin of the conduction electron, and the large $\mathbf{S}$ is for the localized spin. The first scattering term in (6.1) is the $s^{(z)}S^{(z)}$ component. For $s^{(z)}$ we have plus one $(C^{\dagger}_{\mathbf{k}\uparrow} C_{\mathbf{p}\uparrow})$ when the spin of the conduction electron is up, and minus one $(-C^{\dagger}_{\mathbf{k}\downarrow} C_{\mathbf{p}\downarrow})$ when it is down. The other two terms in the scattering interaction represent spin-flip processes, where there is a mutual spin flip between the conduction and local spin. The conduction electrons are assumed to be electrons with spin one-half. The local spin can have any value permitted by quantum mechanics. Here it is taken to have spin S with magnetic quantum number m. Both S and m are either integers or half integers, where m has the range of values $-S \le m \le S$.

$$\begin{aligned} S^{(z)}|m\rangle &= m|m\rangle \\ S^{(+)}|m\rangle &= [S(S+1) - m(m+1)]^{1/2}|m+1\rangle \\ S^{(-)}|m\rangle &= [S(S+1) - m(m-1)]^{1/2}|m-1\rangle \end{aligned} \tag{6.2}$$

The operators $S^{(\pm)}$ raise or lower the magnetic quantum number m. The interaction term in (6.1) can be written compactly as

$$V_{sd} = -\frac{J}{N} \sum_{\mathbf{kp}\alpha\beta} \exp[i\mathbf{R}_j \cdot (\mathbf{k} - \mathbf{p})] \boldsymbol{\sigma}_{\alpha\beta} \cdot \mathbf{S} C^{\dagger}_{\mathbf{k}\alpha} C_{\mathbf{p}\beta} \tag{6.3}$$

The indices α and β denote the spin of the electron after and before the scattering event. The vector $\boldsymbol{\sigma}$ is a Pauli spin matrix for different x, y, and z components. If the meaning of the terms in (6.3) is unclear, just consult (6.1) for the actual formula.

The constant J is called the "exchange energy." Values with $J > 0$ are called "ferromagnetic," since the local spin tends to line up parallel with the conduction band spins. Values with $J < 0$ are called "antiferromagnetic," since the local spin tends to line up antiparallel with the conduction band spins. This nomenclature is dependent upon writing the Hamiltonian with a negative sign before the interaction term. Most magnetic impurities have $J < 0$.

6.1.1. High-Temperature Scattering

The self-energy is calculated for an electron scattering from the impurity. The local spin and the conduction band spins are assumed to have equal probability for any value of m.

These assumptions are valid at high temperature. Whether the temperature is high or low depends upon a reference temperature called the *Kondo temperature* T_K, which will be defined below.

The interaction is (6.3). The average over impurity positions is taken following the prescription in Sec. 4.1.4. The first term is found in first-order perturbation theory. The average of V_{sd} over impurity positions forces $\mathbf{k} = \mathbf{p}$. The first-order self-energy is a constant which is independent of wave vector or energy:

$$\Sigma_1 = -cJ\langle \boldsymbol{\sigma}_{\alpha\alpha} \cdot \mathbf{S}\rangle \tag{6.4}$$

The notation $\boldsymbol{\sigma}_{\alpha\alpha}$ means to take the z component. It is usually assumed that the system is nonmagnetic. Then the average value of $\sigma^{(z)}$ and $S^{(z)}$ are both zero, since there are equal numbers of states occupied with spin up as with spin down. Then the above self-energy expression is zero.

Next consider the second-order scattering. The $n = 2$ term in the S-matrix expansion for the Green's function is

$$\begin{aligned}\mathscr{G}^{(2)}_{\alpha\beta}(\mathbf{k},\tau) &= \tfrac{1}{2}\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \langle T C_{\mathbf{k}\alpha}(\tau)\hat{V}_{sd}(\tau_1)\hat{V}_{sd}(\tau_2)C^\dagger_{\mathbf{k}\beta}(0)\rangle \\ &= -J^2\frac{c}{2}\int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \sum_{ss'uu'}\sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{p}_1\mathbf{p}_2} \boldsymbol{\sigma}^{(\mu)}_{ss'}\boldsymbol{\sigma}^{(\nu)}_{uu'}\langle T S^{(\mu)}(\tau_1)S^{(\nu)}(\tau_2)\rangle \\ &\quad\times \langle T C_{\mathbf{k}\alpha}(\tau)C^\dagger_{\mathbf{k}_1 s}(\tau_1)C_{\mathbf{p}_1 s'}(\tau_1)C^\dagger_{\mathbf{k}_2 u}(\tau_2)C_{\mathbf{p}_2 u'}(\tau_2)C^\dagger_{\mathbf{k}\beta}(0)\rangle\delta(\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{p}_1 + \mathbf{p}_2)\end{aligned} \tag{6.5}$$

The last bracket contains three raising and three lowering operators for conduction electrons. One can pair them up using Wick's theorem. The correlation function of the local spin is not treated as easily. Local spins are neither fermions nor bosons. The thermodynamic average is obtained by averaging over the states $|m\rangle$ with different magnetic quantum number. If there is no external magnetic field, then each state of different m is equally likely, which is assumed here.

The major problem with evaluating the correlation function of the local spin is that they do not obey a Wick's theorem. Here it is no problem, since there are only two operators, which must be paired. However, the lack of Wick's theorem is a major algebraic difficulty in higher orders of perturbation theory. The $S^{(\nu)}$ operators must be paired: each $S^{(+)}$ must be paired with a $S^{(-)}$. Each bracket can have an arbitrary number of $S^{(z)}$ operators, and they are not paired. The correlation function for the local spin has three possible combinations which are evaluated using (6.2):

$$\langle T S^{(+)}(\tau_1)S^{(-)}(\tau_2)\rangle = S(S+1) - m^2 - m\,\mathrm{sgn}(\tau_1 - \tau_2) \tag{6.6}$$

$$\langle T S^{(-)}(\tau_1)S^{(+)}(\tau_2)\rangle = S(S+1) - m^2 + m\,\mathrm{sgn}(\tau_1 - \tau_2) \tag{6.7}$$

$$\langle T S^{(z)}(\tau_1)S^{(z)}(\tau_2)\rangle = m^2 \tag{6.8}$$

The only τ dependence is in the order of the operators. At this point it is useful to average over the values of m. Then the term linear in m vanishes, and the first two of the above expressions become identical.

Assume that the initial spin α of the conduction electron is $\pm$. If the scattering does not flip the spin, the intermediate and final states are also α. Then the local spin correlation function has the form $\langle S^{(z)}S^{(z)}\rangle$. The other possibility is that the spin flips so that the intermediate state has the opposite spin to α. Then one must use one of the two combinations with $S^{(+)}$ and $S^{(-)}$. The Green's functions for the conduction electron are the same for the two

cases of spin-flip or no spin-flip. Just add these two cases, which causes the m^2 terms to cancel. The second-order self-energy for the Kondo model is

$$\Sigma_2(ik_n) = cS(S+1)J^2 \frac{1}{N} \sum_{\mathbf{k}} \frac{1}{ik_n - \varepsilon_{\mathbf{k}}} \tag{6.9}$$

This self-energy expression has the same form as ordinary impurity scattering in the second Born approximation. The spin enters only as an effective interaction of $S(S+1)J^2$. Early workers in the field stopped here, and concluded that spin-flip scattering provided no interesting behavior. This conclusion is wrong, since the unusual effects are found in the higher orders of perturbation theory.

The self-energy of the conduction electron is now calculated in the third Born approximation, which repeats Kondo's original calculation. The anomalous behavior which gives rise to the Kondo effect becomes evident. The $n = 3$ term in the Green's function expansion is

$$\begin{aligned} \mathscr{G}^{(3)}_{\alpha\alpha}(\mathbf{k}, \tau) = -\frac{cJ^3}{N^2} \sum_{\mathbf{pq}} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \int_0^\beta d\tau_3 \mathscr{G}(\mathbf{k}, \tau - \tau_1)\mathscr{G}(\mathbf{p}, \tau_1 - \tau_2) \\ \times \mathscr{G}(\mathbf{q}, \tau_2 - \tau_3)\mathscr{G}(\mathbf{k}, \tau_3) L(\tau_1, \tau_2, \tau_3) \end{aligned} \tag{6.10}$$

$$L(\tau_1, \tau_2, \tau_3) = \sum_{\mu\nu\lambda} \sigma^{(\nu)}_{\alpha s} \sigma^{(\mu)}_{ss'} \sigma^{(\lambda)}_{s'\alpha} \langle T S^{(\nu)}(\tau_1) S^{(\mu)}(\tau_2) S^{(\lambda)}(\tau_3) \rangle \tag{6.11}$$

The self-energy must be the same for spin-up and for spin-down electrons, so set $\alpha = +$ for spin up. The (s, s') can be either of the four combinations $(+, +)$, $(+, -)$, $(-, +)$, $(-, -)$. For $\alpha = +$ then each of these four combinations of (s, s') has a unique choice for $(\nu\mu\lambda)$. Since the electron Green's functions do not depend upon the spin label s, the summation over (s, s') is included in L:

$$\begin{aligned} L(\tau_1, \tau_2, \tau_3) = \langle T S^{(z)} S^{(z)} S^{(z)} \rangle + \langle T S^{(z)}(\tau_1) S^{(-)}(\tau_2) S^{(+)}(\tau_3) \rangle + \langle T S^{(-)}(\tau_1) S^{(+)}(\tau_2) S^{(z)}(\tau_3) \rangle \\ + \langle T S^{(-)}(\tau_1) S^{(z)}(\tau_2) S^{(+)}(\tau_3) \rangle \end{aligned} \tag{6.12}$$

No time variables are written in the first term on the right since this expression is independent of the τ variables. This term equals m^3 for any τ ordering. The other terms depend upon the τ ordering. The factor of $S^{(+)}$ is always to the right of $S^{(-)}$. This choice follows from the ansatz that $\alpha = +$. Spin-flip scattering must lower the spin of the conduction electron, while raising the spin of the local electron. The local raising operator is $S^{(+)}$ which must act at the earlier time. The factors $\sigma_{ss'}$ are not written for each term. They are always plus except for one case. When the factor $S^{(z)}$ immediately follows in time (τ) the operator $S^{(+)}$ then a minus sign is inserted. The minus sign arises because the factor $S^{(+)}$ raises the local spin while flipping the conduction spin from up to down. The $\sigma^{(z)}$ operates on the down conduction spin and gives minus one.

The above expression must be evaluated for all six arrangements of τ ordering. The ordering $\tau_1 > \tau_2 > \tau_3$ gives

$$L = m^3 + [S(S+1) - m^2 - m][m + m - (m+1)] \tag{6.13}$$

$$L = m[S(S+1) + 1] - S(S+1) \tag{6.14}$$

The terms are written in the same order they occur in (6.12). The minus sign before the factor of $(m+1)$ is due to the sign reversal mentioned above. The term proportional to m vanishes when we average over m. After averaging, the above expression equals $-S(S+1)$.

Next consider the τ-ordering of $\tau_3 > \tau_2 > \tau_1$, which gives

$$L = m^3 + [S(S+1) - m^2 + m][m - m + (m+1)] \tag{6.15}$$

$$L = m[S(S+1) + 1] + S(S+1) \tag{6.16}$$

After averaging over m this term gives $S(S+1)$. This result is similar to the one above, except that the sign has changed. One can work out the other four cases. The three cyclic arrangements $\tau_1 > \tau_2 > \tau_3$, $\tau_2 > \tau_3 > \tau_1$, and $\tau_3 > \tau_1 > \tau_2$ each have $L = -S(S+I)$. The other three arrangements $\tau_3 > \tau_2 > \tau_1$, $\tau_2 > \tau_1 > \tau_3$, and $\tau_1 > \tau_3 > \tau_2$ each have $L = S(S+1)$. The three τ integrals in (6.10) must be broken up into the six separate regions of different τ ordering. Three of them are multiplied by $S(S+1)$ while the other three are multiplied by $-S(S+1)$.

The notation (123) means $\tau_1 > \tau_2 > \tau_3$. In this notation the following result is obtained for L after averaging over values of m

$$L = -S(S+1)[(123) + (231) + (312) - (321) - (213) - (132)]$$
$$1 = (123) + (231) + (312) + (321) + (213) + (132)$$

The second line above expresses the idea that the summation of all possible time orderings is just the unity operator. This result can be used to rewrite L as

$$L = -2S(S+1)[(123) + (231) + (312) - \tfrac{1}{2}] \tag{6.17}$$

The amount of work has been cut in half, since now only three time orderings need to be evaluated.

The third-order Green's function of frequency is evaluated by multiplying Eq. (6.10) by $\exp(ik_n\tau)$ and integrating $d\tau$ between $(0, \beta)$. The integrals are most easily evaluated in reverse order:

$$\begin{aligned}\mathscr{G}^{(3)}(\mathbf{k}, ik_n) = -cS(S+1)\frac{J^3}{N^2}\int_0^\beta d\tau_3 e^{\tau_3(ik_n-\varepsilon_k)}\mathscr{G}(\mathbf{k}, \tau_3)\\ \times \int_0^\beta d\tau_2 e^{(ik_n-\varepsilon_p)(\tau_2-\tau_3)}\mathscr{G}(\mathbf{p}, \tau_2-\tau_3)\int_0^\beta d\tau_1 e^{(ik_n-\varepsilon_q)(\tau_1-\tau_2)}\mathscr{G}(\mathbf{q}, \tau_1-\tau_2)\\ \times \int_0^\beta d\tau e^{ik_n-\varepsilon_k)(\tau-\tau_1)}\mathscr{G}(\mathbf{k}, \tau-\tau_1)L(\tau_1, \tau_2, \tau_3)\end{aligned} \tag{6.18}$$

In Eq. (6.17) the last term in brackets is $\frac{1}{2}$. It has no limits on the order of the τ integrals. Then all of the above integrals have their usual limits of $(0, \beta)$, and the above expression is

$$-S(S+1)\frac{cJ^3}{N^2}\mathscr{G}^{(0)}(\mathbf{k}, ik_n)^2\sum_{\mathbf{pq}}\frac{1}{(ik_n-\varepsilon_p)(ik_n-\varepsilon_q)} \tag{6.19}$$

This expression is a contribution to the third-order Green's function. If one eliminates the prefactor of $\mathscr{G}^{(0)}(\mathbf{k}, ik_n)^2$ then the remaining expression is a contribution to the third-order self-energy. It is just the third term in the T-matrix equation generated by iterating (4.49). This contribution to the self-energy is rather dull.

The Kondo effect comes from the terms in L which restrict the order of the τ integrals. The restrictions are on τ_1, τ_2 and τ_3. There is no restriction on the $d\tau$ integral. It can be done

immediately, and yields $\mathcal{G}^{(0)}(\mathbf{k}, ik_n)$. The first interesting integral is (123), whose limits can be written as

$$(123) \Rightarrow \int_0^\beta d\tau_3 \int_{\tau_3}^\beta d\tau_2 \int_{\tau_2}^\beta d\tau_1 \tag{6.20}$$

Doing the integral in (6.18) with these limits gives the following result. Drop all of the prefactors and summations, and list only the results of the three τ integrals:

$$\begin{aligned}(123) = &\frac{[1-n(p)][1-n(q)]}{(ik_n-\varepsilon_k)(ik_n-\varepsilon_p)(ik_n-\varepsilon_q)} + \frac{[1-n(q)][n(k)-n(p)]}{(ik_n-\varepsilon_p)(\varepsilon_k-\varepsilon_p)(\varepsilon_p-\varepsilon_q)} \\ &+ \frac{[1-n(p)][1-n(q)]}{(ik_n-\varepsilon_q)(\varepsilon_k-\varepsilon_q)(\varepsilon_q-\varepsilon_p)}\end{aligned} \tag{6.21}$$

The expression has a multitude of fermion occupation factors $n(k)$, etc. Similar expressions are obtained for the other two τ orderings (231) and (312):

$$(231) \Rightarrow \int_0^\beta d\tau_3 \int_{\tau_3}^\beta d\tau_2 \int_0^{\tau_2} d\tau_1$$

$$\begin{aligned}(231) = &\frac{1}{(ik_n-\varepsilon_p)(\varepsilon_p-\varepsilon_q)}\left\{\frac{[1-n(q)][n(k)-n(p)]}{\varepsilon_p-\varepsilon_k} - \frac{[1-n(p)][n(k)-n(q)]}{\varepsilon_q-\varepsilon_k}\right. \\ &\left. + \frac{n(p)[1-n(q)]}{ik_n-\varepsilon_k} + \frac{n(k)n(p)[1-n(q)]-n(q)[1-n(k)][1-n(p)]}{ik_n-\varepsilon_k-\varepsilon_k-\varepsilon_p+\varepsilon_q}\right\}\end{aligned} \tag{6.22}$$

$$(312) \Rightarrow \int_0^\beta d\tau_3 \int_0^{\tau_3} d\tau_2 \int_{\tau_2}^{\tau_3} d\tau_1$$

$$\begin{aligned}(312) = &\frac{[1-n(p)][n(k)-n(q)]}{(ik_n-\varepsilon_p)(ik_n-\varepsilon_q)(\varepsilon_p-\varepsilon_k)} - \frac{n(q)[1-n(p)]}{(ik_n-\varepsilon_k)(ik_n-\varepsilon_q)(\varepsilon_p-\varepsilon_q)} \\ &- \frac{n(k)n(p)[1-n(q)]-n(q)[1-n(k)][1-n(p)]}{(ik_n-\varepsilon_p)(ik_n-\varepsilon_k-\varepsilon_p+\varepsilon_q)(\varepsilon_p-\varepsilon_q)}\end{aligned} \tag{6.23}$$

An interesting thing happens when we add these three expressions. Most of the terms cancel:

$$(123)+(231)+(312) =$$

$$-\frac{1}{(ik_n-\varepsilon_k)^2}\left[\frac{1}{(ik_n-\varepsilon_p)(ik_n-\varepsilon_q)} + \frac{n(p)}{(ik_n-\varepsilon_q)(\varepsilon_q-\varepsilon_p)} + \frac{n(q)}{(ik_n-\varepsilon_p)(\varepsilon_p-\varepsilon_q)}\right] \tag{6.24}$$

A further simplification occurs by noticing that the last two terms are identical after interchanging the variables of integration p and q. The result for the third-order self-energy is

$$\Sigma_3(ik_n) = -S(S+1)\frac{cJ^3}{N^2}\sum_{\mathbf{pq}}\frac{1}{(ik_n-\varepsilon_q)}\left[\frac{1}{(ik_n-\varepsilon_p)} - \frac{4(n(p)}{\varepsilon_p-\varepsilon_q}\right] \tag{6.25}$$

The final expression is rather simple. There are two terms. The first term in the brackets is the rather dull term, which is the third-order Born scattering from a simple potential. The second term in the brackets was first derived by Kondo. He used it to explain the temperature dependence of the resistivity, which is now called the Kondo effect. Today it is known that similar anomalous terms occur in all higher orders of perturbation theory. The modern theory

of the Kondo effect employs terms from all orders of perturbation theory. Nevertheless, Kondo's simple arguments will be presented.

The self-energy expression has a factor of the fermion occupation number $n(p)$. These factors do not occur in scattering from a simple potential. They occur here because of the spin-flip scattering. Interesting physics happens whenever they occur, such as in the BCS theory of superconductivity or in the edge singularities in X-ray absorption in metals.

The second term in the above brackets produces a term in the resistivity that goes as $\ln(T)$, where T is the temperature. A quick and sloppy derivation is presented of this behavior. The conduction band density of states is defined as $g(\varepsilon)$. It is assumed to be a smooth function of energy ε, which extends from $W < \varepsilon < B$. Summations over wave vector are changed to integrations over energy using the prescription

$$\frac{1}{N}\sum_{\mathbf{p}} f(\varepsilon_{\mathbf{p}}) \Rightarrow \int_{-W}^{B} d\varepsilon g(\varepsilon) f(\varepsilon) \tag{6.26}$$

A typical alloy system that shows the Kondo effect is manganese impurities in copper. The density of states $g(\varepsilon)$ of conduction electrons in metallic copper has a large energy dependence because of the occupied d-bands. However, the conduction electrons that are involved in the Kondo effect have energies within several $k_B T$ of the Fermi energy. On that small energy scale, the density of states of copper and most metals is smooth and featureless. The custom is to call the density of states $g(\varepsilon)$ a constant $g(0)$, where $\varepsilon = 0$ signifies the Fermi energy.

For the Kondo effect, the interesting third-order self-energy is

$$\Sigma_{3b}(ik_n) = -4S(S+1)cJ^3 \int_{-W}^{B} d\varepsilon_1 \frac{g(\varepsilon_1)}{ik_n - \varepsilon_1} \int_{-W}^{B} d\varepsilon_2 \frac{g(\varepsilon_2)n(\varepsilon_2)}{\varepsilon_1 - \varepsilon_2} \tag{6.27}$$

The ultimate goal is to calculate the resistivity. Then the quantity of interest is the scattering rate, which is related to the imaginary part of the self-energy. In this expression let $ik_n \to \omega + i\eta$, where η is infinitesimal, and then take the imaginary part. The first energy denominator gets replaced by $-\pi\delta(\omega - \varepsilon_1)$. The first integral gives $-\pi g(\omega)$. In order to evaluate the second integral, it is assumed that the temperature is low, so the fermion occupation factor can be replaced by the step function $n(\varepsilon_2) \approx \Theta(-\varepsilon_2)$. The most important term is the logarithmic singularity which comes from this Fermi cutoff

$$\int_{-W}^{B} d\varepsilon_2 \frac{g(\varepsilon_2)n(\varepsilon_2)}{\omega - \varepsilon_2} \approx \int_{-W}^{0} d\varepsilon_2 \frac{g(\varepsilon_2)}{\omega - \varepsilon_2} \approx g(\omega) \ln\left|\frac{\omega}{W}\right| \tag{6.28}$$

Since the density of states $g(\omega)$ is a smooth function of energy ω, this quantity is usually replaced by $g(0)$. The standard expression for the third-order self-energy in the Kondo effect is

$$\mathrm{Im}[\Sigma_{3b}(\omega)] = -4\pi S(S+1)cJ^3 g(0)^2 \ln\left|\frac{\omega}{W}\right| \tag{6.29}$$

The real part of Σ_{3b} has an uninteresting ω dependence near $\omega = 0$.

This result is combined with the larger term from the second order of perturbation theory:

$$\mathrm{Im}[\Sigma(\omega)] = -\pi S(S+1)cJ^2 g(0)\left[1 + 4Jg(0)\ln\left|\frac{\omega}{W}\right| \cdots\right] \tag{6.30}$$

In calculating the conductivity, the energy ω of the electron is averaged over a region near the Fermi surface of width k_BT. This replaces $\ln|\omega/W|$ by $\ln(k_BT/W)$. Kondo derived a temperature-dependent resistivity of the form

$$\rho(T) = \rho(0)[1 + 4Jg(0)\ln(k_BT/W)] + bT^5 \tag{6.31}$$

The first term $\rho(T = 0)$ is the temperature-independent part of the impurity scattering. The second term with $\ln(T)$ is the Kondo effect. The last term $\sim T^5$ is the low-temperature contribution from electron–phonon scattering. Usually the experiments are done in metals or alloys where phonon drag does not eliminate the T^5 law.

The factor of $\ln(k_BT/W)$ is negative at low temperature. If $J < 0$ then the Kondo term is positive. At very low temperatures, the resistivity increases in value, since this term becomes large. This expression predicts a minimum in the resistivity, which is observed in Kondo systems. Define $T_{\min}$ as the temperature at the minimum resistance. By taking $d\rho/dT = 0$ one finds that $(T_{\min})^5 \sim c$, where c is the concentration of impurities (Note $\rho(0) \propto c$). This relationship is in agreement with experiments on Kondo alloys. A typical measurement of $\rho(T)$ is shown in Fig. 6.1 for the system of $La_{1-x}Ce_xB_6$. The resistance minimum is evident. The resistance increases at low temperature, and finally saturates. The reason for the saturation is given in the following section.

The theory presented in this section is a high-temperature theory. A key assumption in the derivation is that the local spin has equal energy whether it is up or down. At low temperature, this assumption ceases to be valid. Each local spin becomes locked into a collective state with the conduction band spins. This collective state has a binding energy, which must be overcome during a spin-flip. The spin-flips become frozen out at low temperature, and the Kondo effect saturates. This process is described in the next section. Kondo's formula (6.31) is only valid at temperatures high above the formation of the collective state.

Our derivation of the Kondo resistance was sloppy for the following reasons. In evaluating the integral over $d\varepsilon_2$ the Fermi factor $n(\varepsilon_2)$ was replaced by a step function $\Theta(-\varepsilon_2)$. This approximation is only valid at zero temperature. Later in the derivation, we averaged $\ln|\omega/W|$ over the distribution of electrons while assuming the temperature is nonzero. A better derivation would have evaluated the integral over $d\varepsilon_2$ at nonzero temperature, which broadens the Fermi distribution a few degrees in energy. Of course, this realism makes the integral harder to evaluate.

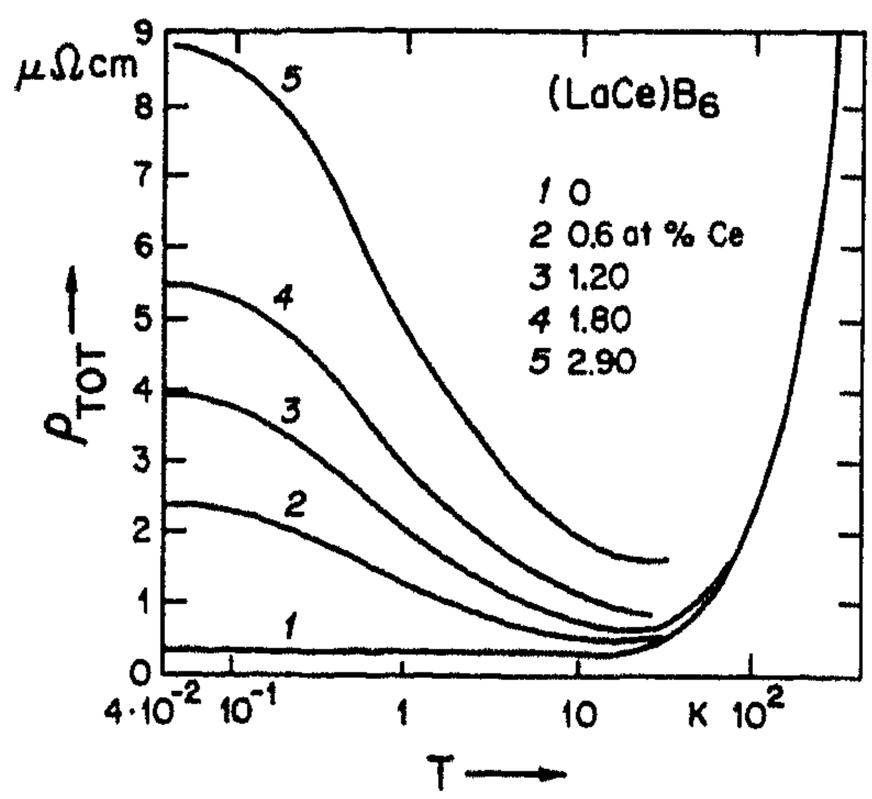


FIGURE 6.1 Total resistivity vs T for LaB_6 and $La_{1-x}Ce_xB_6$. The Kondo temperature is $T_K = 1.05$ K. *Source* Samwer and Winzer (1976) (used with permission).

6.1.2. Low-Temperature State

At low temperature the local spin forms a collective state with the conduction band spins. The collective state is formed from a linear combination of conduction band states, and is a many-particle entity. Earlier chapters have often discussed the role of screening charge in reducing the long range Coulomb interaction in conducting systems. The screening charge is not from a single electron in the conduction band, but usually is formed by a linear combination of conduction states. The present collective state can be viewed, in a similar way, as a screening effect. However, now it is screening the spin of the local state, rather than its charge. If the impurity is charged there could coexist the screening clouds for spin and charge.

For $J < 0$ the state is a singlet. In order to keep the discussion simple, let us assume that the coupling is antiferromagnetic ($J < 0$) and that the local spin has $S = \frac{1}{2}$. The up and down states of the local spin are denoted as α and β. Yosida (1966) first considered the collective state. Its wave function can be written in two possible ways:

$$\psi_{0a} = \frac{1}{\sqrt{N}} \sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}[\alpha C^\dagger_{\mathbf{k}\downarrow} - \beta C^\dagger_{\mathbf{k}\uparrow}]|F\rangle \tag{6.32}$$

$$\psi_{0b} = \frac{1}{\sqrt{N}} \sum_{|\mathbf{k}|<k_F} b_{\mathbf{k}}[\alpha C_{\mathbf{k}\uparrow} + \beta C_{\mathbf{k}\downarrow}]|F\rangle \tag{6.33}$$

$$|F\rangle = \Pi_{|p|<k_F} C^\dagger_{\mathbf{p}\uparrow} C^\dagger_{\mathbf{p}\downarrow}|0\rangle \tag{6.34}$$

The symbol $|F\rangle$ denotes the filled Fermi sea of electrons. The state ψ_{0a} forms the collective state by adding an electron above the Fermi sea. It has the usual form for a spin singlet, as an antisymmetric combination of up and down states of the local and conduction spins. The coefficient $a_{\mathbf{k}}$ is a parameter that needs to be determined.

The second choice of collective wave function ψ_{0b} needs further explanation. Here one is forming the collective state by taking linear combinations of hole states, where the hole is a missing electron from the filled Fermi sea. The easiest way to understand this wave function is to examine one of its terms, which include the relevent factor from the Fermi sea:

$$[\alpha C_{\mathbf{k}\uparrow} + \beta C^\dagger_{\mathbf{k}\downarrow}]C^\dagger_{\mathbf{k}\uparrow} C^\dagger_{\mathbf{k}\downarrow}|0\rangle = [\alpha C^\dagger_{\mathbf{k}\downarrow} - \beta C^\dagger_{\mathbf{k}\uparrow}]|0\rangle \tag{6.35}$$

The right-hand expression is obviously a spin singlet, composed of an antisymmetric combination of down and up spin states. The relative plus sign on the left becomes a minus sign on the right. The two states ψ_{0a} and ψ_{0b} are orthogonal. Each describes a different collective state, and the system will choose the one with the lowest energy. One is a ground state, and one is an excited state.

The first term in (6.1) is called H_0 and the second term is V_{sd}. If there is no local spin, the ground state energy for this model Hamiltonian is

$$H_0|F\rangle = E_G|F\rangle \tag{6.36}$$

$$E_G = 2 \sum_{|\mathbf{k}|<k_F} \varepsilon_{\mathbf{k}} \tag{6.37}$$

Now add one local spin at the point $\mathbf{R}_j = 0$. In this calculation, one does *not* average over the spatial positions of the impurity, and one does not set $\mathbf{p} = \mathbf{k}$ in the interaction term. Only one impurity site is being considered.

The goal is to calculate the additional energy $\delta E < 0$ gained by the system because of the formation of the collective spin state. It is analogous to the energy found for charge screening using Fumi's theorem in Sec. 4.1. The eigenvalue equation is

$$\{H_0 + V_{sd}\}\psi_{0a,b} = (E_G + \delta E_{a,b})\psi_{0,a,b} \tag{6.38}$$

This equation is not strictly valid since neither ψ_{0a} nor ψ_{0b} is an exact eigenstate of the Hamiltonian. In order to understand this assertion, consider the effect of one term in V_{sd} acting upon one term in ψ_{0a}

$$\sum_{\mathbf{p},\mathbf{q}} C^\dagger_{\mathbf{p}\uparrow} C_{\mathbf{q}\downarrow} S^{(-)} \sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}} \alpha C^\dagger_{\mathbf{k}\downarrow} |F\rangle \tag{6.39}$$

The spin-lowering operator $S^{(-)}$ changes α to β. If $\mathbf{q} = \mathbf{k}$ there is a term

$$\left[\sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}\right]\left[\sum_{|\mathbf{p}|>k_l} \beta C^\dagger_{\mathbf{p}\uparrow}|F\rangle\right] \tag{6.40}$$

which is recognized as being similar to one of the other terms in ψ_{0a}. However, there are also terms in (6.39) in which $\mathbf{k}$, $\mathbf{p}$, and $\mathbf{q}$ are all different wave vectors. These terms describe a state with one hole and two excitations above the Fermi surface. An electron–hole pair has been excited in addition to the original excitation. This state is perfectly valid, since it is generated by the interaction term in the Hamiltonian. Our failure to include it in the ansatz eigenstate ψ_{0a} is simply an approximation. A more accurate ansatz would include these additional terms. There are an infinite number of them, which correspond to the multiple excitation of electron–hole pairs. The more pairs that are included, the more accurate is the wave function.

The best way to understand the ansatz wave function in (6.32) is to consider it a variational calculation on the ground state. The coefficients $a_{\mathbf{k}}$ and $b_{\mathbf{k}}$ are the variational parameters. The ground state energy is calculated, using Dirac notation, with $j = a$ or b, as

$$E_G + \delta E_j = \frac{\langle\psi_{0j}|H|\psi_{0j}\rangle}{\langle\psi_{0j}|\psi_{0j}\rangle} \tag{6.41}$$

After evaluating the right-hand side of this expression, say for $j = a$, one finds the minimum value of δE_a by taking the functional derivative $\delta(\delta E_a)/\delta a_{\mathbf{k}}$. There results an equation for $a_{\mathbf{k}}$ that can be solved. These steps produce the minimum ground state energy for the ansatz (6.32).

The method of solution is equivalent to the variational procedure, but is algebraically simpler. Equation (6.38) is evaluated. The multipair states are neglected, and only project out the terms with one conduction band excitation. For example, when solving for ψ_{0a} define $\langle F|\alpha C_{\mathbf{k}\downarrow}$ as the Hermitian conjugate of $\alpha C^\dagger_{\mathbf{k}\downarrow}|F\rangle$ and evaluate the expression

$$\langle F|\alpha C_{\mathbf{k}\downarrow}\{H_0 + V_{sd} - E_G - \delta E_a\}\psi_{0a} = 0 \tag{6.42}$$

As an example, the easy terms are done first:

$$\{H_0 - E_G - \delta E_a\}\psi_{0a} = \frac{1}{\sqrt{N}} \sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}[E_G + \varepsilon_{\mathbf{k}} - E_G - \delta E_a][\alpha C^\dagger_{\mathbf{k}\downarrow} - \beta C^\dagger_{\mathbf{k}\uparrow}]|F\rangle$$

$$\langle F|\alpha C_{\mathbf{k}\downarrow}\{H_0 - E_G - \delta E_a\}\psi_{0a} = \frac{1}{\sqrt{N}} a_{\mathbf{k}}[\varepsilon_{\mathbf{k}} - \delta E_a] \tag{6.43}$$

A similar evaluation must be done for V_{sd}. One term was previously evaluated in (6.39). Another term in $V_{sd}\psi_{0a}$ is

$$\sum_{\mathbf{p},\mathbf{q}} S^{(z)}(C^\dagger_{\mathbf{p}\uparrow}C_{\mathbf{q}\uparrow} - C^\dagger_{\mathbf{p}\downarrow}C_{\mathbf{q}\downarrow}) \sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}\alpha C^\dagger_{\mathbf{k}\downarrow}]|F\rangle \tag{6.44}$$

The operator $S^{(z)}$ acts upon α and gives $\alpha/2$. The electron operators give several contributions. For $\mathbf{p} = \mathbf{q}$ the part in parentheses gives the number of electrons with spin up minus the number with spin down. The state $C^\dagger_{\mathbf{k}\downarrow}|F\rangle$ has one more electron with spin down, so the summation gives -1. There is also a term where $\mathbf{q} = \mathbf{k}$. These two contributions are

$$-\tfrac{1}{2}\sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}\alpha C^\dagger_{\mathbf{k}\downarrow}|F\rangle - \frac{1}{2}\left[\sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}\right]\left[\sum_{|\mathbf{p}|>k_F} \alpha C^\dagger_{\mathbf{p}\downarrow}|F\rangle\right] \tag{6.45}$$

Other terms contain an electron–hole pair, and these contributions are being ignored. In the above expression, the first term can be neglected since it is of order $O(1/N)$ compared to the second. The second has a double summation, while the first has only a single summation. By employing these kinds of approximations, there results

$$V_{sd}\psi_{0a} = \frac{3J}{2N^{3/2}}\left[\sum_{|\mathbf{k}|>k_F} a_{\mathbf{k}}\right]\sum_{|\mathbf{p}|>k_F} a_{\mathbf{k}}(\alpha C^\dagger_{\mathbf{p}\downarrow} - \beta C^\dagger_{\mathbf{p}\uparrow})|F\rangle \tag{6.46}$$

This result is combined with (6.43) to produce the equation for $a_{\mathbf{k}}$:

$$a_{\mathbf{k}} = -\frac{3J}{2N}\frac{1}{\varepsilon_{\mathbf{k}} - \delta E_a}\sum_{|\mathbf{p}|>k_F} a_{\mathbf{p}} \tag{6.47}$$

This expression is then summed over $\mathbf{k}$. These summations are changed to integrals over the particle energy. The eigenvalue equation is

$$1 = -\frac{3J}{2}\int_0^B d\varepsilon \frac{g(\varepsilon)}{\varepsilon - \delta E_a} \tag{6.48}$$

The solution to the wave function $\psi_{\sigma b}$ follows the same steps used to obtain ψ_{0a}. The form of the wave function in (6.35) has the same form as (6.32), so that all steps in the derivation are similar. There are two minor differences. First, since $|\mathbf{k}| < k_F$, the integral over energy is over negative energy states. Second, the result of $H_0\psi_{0b}$ has the factor of $(E_G - \varepsilon_{\mathbf{k}})$; the minus sign preceding $\varepsilon_{\mathbf{k}}$ comes from the fact that ψ_{0b} has an electron missing from the Fermi sea. This changes the sign of ε in the integral. So the eigenvalue equation for ψ_{0b} has the form

$$b_{\mathbf{k}} = \frac{3J}{2N}\frac{1}{\varepsilon_{\mathbf{k}} + \delta E_b}\sum_{|\mathbf{p}|<k_F} b_{\mathbf{p}} \tag{6.49}$$

$$1 = \frac{3J}{2}\int_{-W}^0 d\varepsilon \frac{g(\varepsilon)}{\varepsilon + \delta E_b} \tag{6.50}$$

The two eigenvalue equations (6.48) and (6.50) are very similar. The major difference is that one has an integral over the empty states in the conduction band, while the other has an integral over the occupied states. Since $\delta E_{a,b} < 0$ neither energy denominator is singular.

These two expressions will be evaluated for a simple model of the conduction band. The density of states $g(\varepsilon)$ equals a constant g. Then the two integrals give

$$1 = (3Jg/2)\ln\left|\frac{\delta E_a}{W - \delta E_a}\right| \tag{6.51}$$

$$1 = (3Jg/2)\ln\left|\frac{\delta E_b}{B - \delta E_b}\right| \tag{6.52}$$

which have the solution

$$\delta E_a = -\frac{W\Lambda}{1+\Lambda} \tag{6.53}$$

$$\delta E_b = -\frac{B\Lambda}{1+\Lambda} \tag{6.54}$$

$$\ln(\Lambda) = \frac{2}{3Jg} \tag{6.55}$$

The largest value of $\delta E_{a,b}$ is determined by whether B or W is largest. If the band is less than half full, then $W < B$, $\delta E_a > \delta E_b$, and ψ_{0b} is the ground state. If the band is more than half full, then $W > B$ and ψ_{0a} is the ground state. If the band is exactly half full, the states are degenerate. These results only apply to the case of a constant density of states. Other examples are assigned in the problems.

Remember it is assumed that $J < 0$. In fact, there is no solution for singlet states when $J > 0$. The present approximations only make sense when $\delta E \ll (B, W)$ so that the logarithm terms in (6.52) are negative. Then the equation has a solution only for $J < 0$. For antiferromagnetic coupling, the factor of $\Lambda = \exp(2/3Jg)$ is generally much less than unity. This expression also has the feature that the coupling constant J appears as an inverse power in an exponent. It would be impossible, or at least very difficult, to derive an expression for δE from perturbation theory.

This mathematical form, of an inverse power in an exponent, will be encountered in Chap. 10 in the BCS theory of superconductivity. The energy gap and the formula for T_c have a similar factor of $\exp[-1/N_F V_0]$.

The *Kondo temperature* is defined by the expression

$$k_B T_K = W \exp\left\{-\frac{1}{2g(0)|J|}\right\} \tag{6.56}$$

This expression is similar to the result for the change in ground state energy. The main difference is the factor of $\frac{2}{3}$ in the exponent has become $\frac{1}{2}$. The reason for this change is given below. A general definition is that the change in ground state energy defines an energy scale that also defines a characteristic temperature. The Kondo temperature is also very close to the temperature at which the Kondo effect becomes important in the resistivity. The self-energy correction given by the second term in (6.30) becomes equal to the first at a temperature of $T \sim T_K$. It is not the same temperature as the resistance minimum, since the minimum depends upon the concentration of impurities. The Kondo temperature defines an energy scale for a single impurity. Values of T_K for actual alloys span a large range in temperature, from 10^{-3} to 10^3 K.

So far the derivation suggests that the collective state has a similar behavior regardless of the sign of J. However, further work has shown that the collective state is not formed for ferromagnetic coupling ($J > 0$). If further terms are evaluated in the perturbation expansion,

they destabilize the collective state for $J > 0$. For $J < 0$, further terms in the perturbation theory continue to predict a collective state. The Kondo effect does depend upon the sign of J, and exists only for $J < 0$.

6.1.3. Kondo Temperature

The two prior sections discussed the high- and low-temperature regimes of the Kondo effect. The most interesting behavior occurs for temperatures near the Kondo temperature T_K, which is defined in (6.56). A short and somewhat qualitative discussion of this temperature region is presented here. This region is actually very well understood because the exact solution to the Kondo problem is now known. See the solutions by Andrei *et al.* (1983) and Wiegmann (1981).

The high-temperature scattering theory has been carried out to much higher orders of perturbation theory. Abrikosov (1965), and Silverstein and Duke (1967) have suggested that the perturbation expansion for the retarded Green's function can be represented by the simple function

$$\Sigma_{\text{ret}}(\omega) = \frac{cJ}{1 - 2Jg(\omega)\ln|\omega/W| - i\pi Jg(\omega)S(S+1)} \tag{6.57}$$

The function calculated in Sec. 6.1.1 was the imaginary part of the self-energy. The imaginary part of the above expression is

$$\text{Im}[\Sigma_{\text{ret}}(\omega)] = \frac{\pi c J^2 g(\omega)S(S+1)}{[1 - 2Jg(\omega)\ln|\omega/W|]^2 + [\pi Jg(\omega)S(S+1)]^2} \tag{6.58}$$

If the denominator is expanded in a power series in $Jg(\omega)$, the first term is the perturbation expression in (6.30). Another feature of this expression is that the denominator has a resonance when the energy $\omega \sim W\exp[1/2Jg(0)]$, which makes sense when $J < 0$. There is no resonance for $J > 0$. This observation provides further evidence that there is no collective behavior for $J > 0$.

A factor of $2Jg(\omega)$ can be taken out of the denominator in (6.57). Using the definition of the Kondo temperature in (6.56) allows the real part of the denominator to have a simple form:

$$\begin{aligned} \Sigma_{\text{ret}}(\omega) &= -\frac{c}{2g(\omega)}\frac{1}{\ln|\omega/k_BT_K| + i\pi S(S+1)/2} \\ \text{Im}[\Sigma_{\text{ret}}(\omega)] &= \frac{c\pi S(S+1)/4g(\omega)}{\ln^2|\omega/k_BT_K| + [\pi S(S+1)/2]^2} \end{aligned} \tag{6.59}$$

The interesting feature of this expression is that the ω dependence of the self-energy scales only with the Kondo temperature. One should be able to describe all of the self-energy effects by this single parameter. It also has a resonance behavior. The square of the logarithm has its smallest value of zero at $\omega_K = k_BT_K$, and the denominator increases in magnitude for values of ω different from ω_K.

6.1.4. Kondo Resonance

Wilkins (1982) observed that many of the properties associated with the Kondo effect can be described by a simple expression. The system acts as if the density of states has a

resonance at the Fermi surface for each impurity with a local moment. Let the concentration of impurities be $c_i = N_i/N$. A simple expression which well describes the density of states is

$$g_K(\omega) = g(\omega) + \frac{c_i}{\pi}\frac{\gamma}{\omega^2 + \gamma^2} = g(\omega) + \delta g(\omega) \tag{6.60}$$

where $\gamma = 1.6k_B T_K$. The change $\delta g(\omega)$ integrates to one electron for each impurity with a local spin.

For example, to calculate the change in the specific heat δC due to the change $\delta g(\omega)$, first find the change δU in the internal energy

$$\begin{aligned} \delta C &= \frac{\partial}{\partial T}\delta U = c_i f(T/T_K) \\ \delta U &= \int d\omega \omega n_F(\omega)\delta g(\omega) \end{aligned} \tag{6.61}$$

Figure 6.2 shows how the function $f(T/T_K)$ (dashed line) compares to the exact result (solid line). The two results are quite similar. Note that the temperature scale spans several decades. The specific heat per magnetic impurity is a function only of T/T_K. At high temperature the specific heat per impurity goes as $f \sim T_K/T$. At low temperatures it goes as $f \sim T/T_K$, which can be large if the Kondo temperature is low. Wilkins observes that the entire effect of the many-body interaction is to add a *Kondo Resonance* at the Fermi surface, such as the term $\delta g(\omega)$. Numerical calculations of the Kondo resonance show that it is usually asymmetric in frequency and is not centered precisely at $\omega = 0$.

The resonance behavior also affects the resistivity. Assume that the Kondo resonance affects the quasiparticle self-energy $\Gamma = \mathrm{Im}[\Sigma]$. Instead of the behavior predicted in (6.59) use the resonance form in (6.60). The temperature-dependent conductivity $\sigma(T)$ is given in terms of the lifetime $\tau(T)$

$$\sigma = \frac{n_0 e^2 \tau}{m} \tag{6.62}$$

$$\tau = \int_{-\infty}^{\infty} d\omega \left[-\frac{dn_F(\omega)}{d\omega}\right]\frac{1}{2\Gamma} \tag{6.63}$$

$$2\Gamma(\omega) = \left(1 + \frac{r\gamma}{\omega^2 + \gamma^2}\right)/\tau_0 \tag{6.64}$$

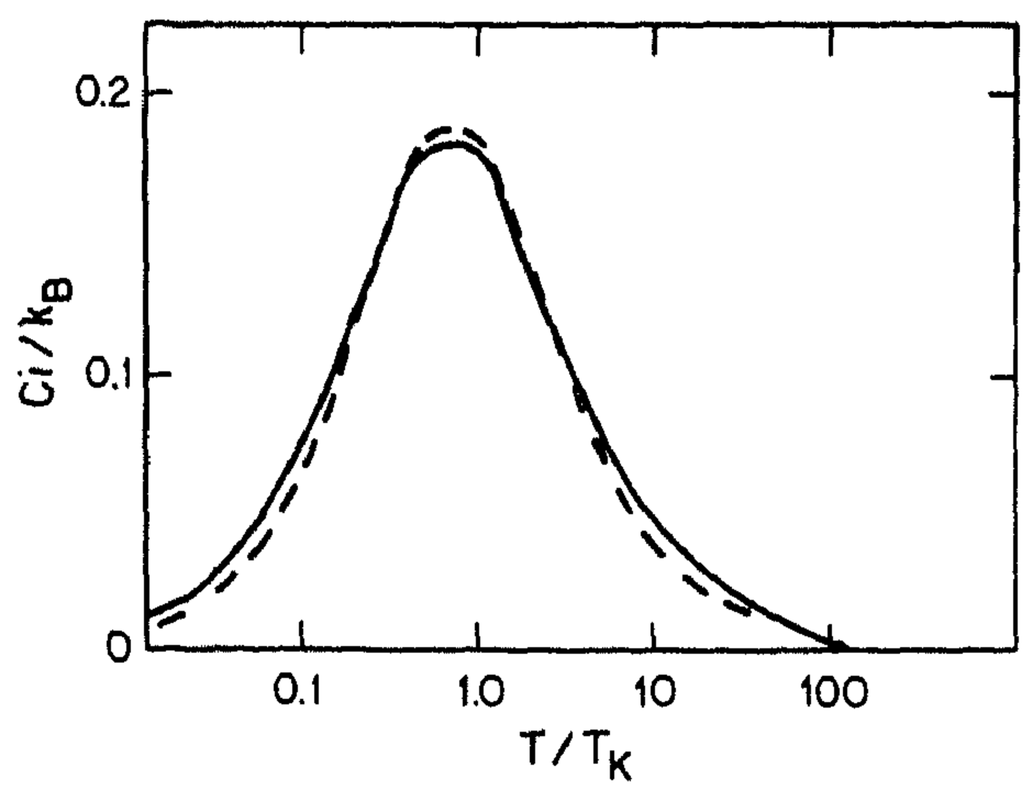


FIGURE 6.2 Specific heat per impurity for the Kondo model. The solid line is the numerical solution using the renormalization group. The dashed line is the calculation using the Lorentzian resonance with unit weight and width equal to $1.6T_K$. Adapted from Wilkins (1982) (used with permission).

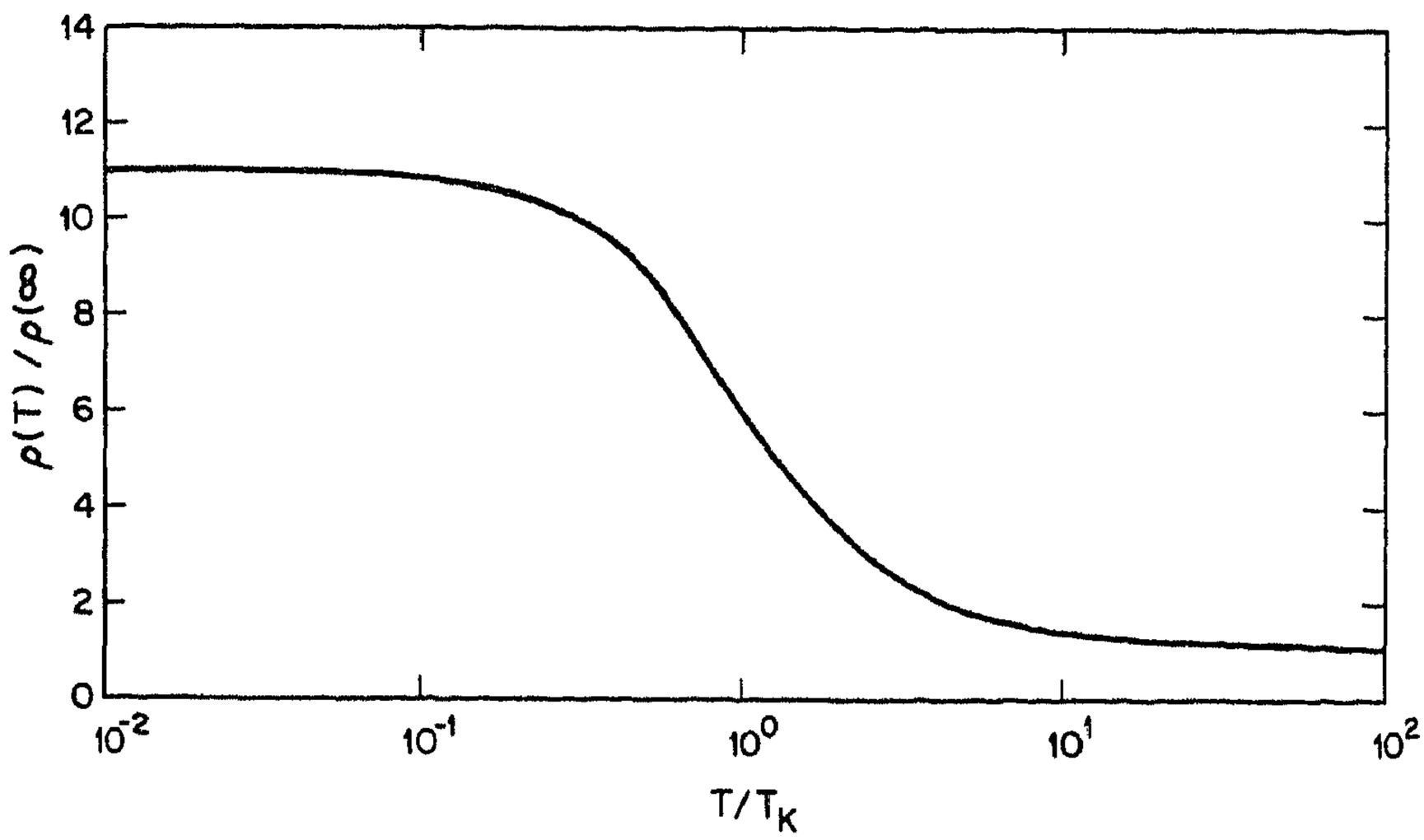


FIGURE 6.3 Resistivity vs temperature in the Kondo model.

where $\sigma_0 = n_0 e^2 \tau_0 / m$ is the conductivity from nonresonant impurity scattering for $T \gg T_K$. The parameter r comes from the ratio of resonant to nonresonant impurity scattering. The frequency integral in (6.63) can be done numerically. The result is given in Fig. 6.3, for $r = 10$, which shows the resistivity $\rho(T) = 1/\sigma(T)$. At high temperature ($T \gg T_K$) one finds $\rho = \rho_0$. At low temperature ($T \ll T_K$) one finds that $\rho = \rho_0(1 + r)$. This behavior of the resistivity is similar to the experimental dependence shown in Fig. 6.1. At high temperature the resonance effects from spin-flip scattering are smeared out by thermal broadening. As the temperature is lowered to the vicinity of T_K, the thermal smearing decreases, and the Kondo resonance becomes more important in scattering the electron. The resistivity increases. At still lower temperatures, the collective state begins to form, and the local spin becomes harder to flip. Then the resistance saturates, and no longer increases with decreasing temperature.

6.2. SINGLE-SITE ANDERSON MODEL

The Anderson model (1961) was introduced in Sec. 1.4. It is another model for a system of conduction electrons that interact with a local spin. Equation (1.354) gives the transformation of Schrieffer and Wolff (1966), which shows that the model has some terms that are similar to those of the Kondo model. Early workers thought that the two models made very similar predictions. Now it is known that the Anderson model has a greater variety of behavior. It has the more interesting physics.

The Kondo model treats the local spin as a separate entity. The Anderson model treats the local spin as just another electron. It can undergo exchange and other processes with the conduction electrons. The Anderson model is more realistic.

The Anderson model is written here with a slightly altered notation. Most applications of the model consider the localized state to have spin and orbital degeneracy. It could be a d electron of 10 states or an f electron of 14 states. Usually the spin–orbit interaction splits these states into 6 and 4, and into 8 and 6, respectively. The crystal field may cause further splittings. Let N_f denote the actual degeneracy of the local level: it may be an even number such as 2, 4, 6, 8, 10, or 14. One of the theoretical developments is a perturbation expansion, where the expansion parameter is $1/N_f$. Obviously this expansion works better for larger

values of N_f. Many recent applications of the Anderson model have been for *heavy fermion* systems, where the local orbital is one or several electrons in f orbitals. That is the origin of the subscript f on the factor of N_f. The symbol N without the subscript is the number of atoms sites in the solid. For reviews of heavy fermions see Stewart (1984) or Lee *et al.* (1986).

The Hamiltonian is written as

$$H = H_0 + V \tag{6.65}$$

$$H_0 = \sum_{k\nu} \varepsilon_k C^\dagger_{k\nu} C_{k\nu} + \varepsilon_f \sum_\mu n_\mu + U \sum_{\mu>\nu} n_\mu n_\nu \tag{6.66}$$

$$n_\mu = f^\dagger_\mu f_\mu \tag{6.67}$$

$$V = \frac{1}{\sqrt{N}} \sum_{k\nu} V_{k\nu} [f^\dagger_\nu C_{k\nu} + C^\dagger_{k\nu} f_\nu] \tag{6.68}$$

where V is the hybridization interaction between the band electrons (k) and the localized electrons f_μ. The onsite Coulomb repulsion is U and ε_f is the eigenvalue for a single f-electron. The summations over (μ, ν) run between 1 and N_f. This index denotes a combination of orbital and spin states. Only the band states (k, ν) with the same index interact with the local orbitals. The wave vector **k** can be represented by the magnitude k plus various spin and angular momentum states. The latter are represented by the index ν. An important point is that of all the many symmetries available to the conduction band states, the only ones which couple to the local state are those values of ν between 1 and N_f. The summations over k will eventually be changed to an integral over the energy ε. The density of states is $g(\varepsilon)$ and the matrix element as a function of energy is written as $V_k = V(\varepsilon)$.

How does the local spin get flipped in the Anderson model? Actually it does not flip. Because of the hybridization term, a local spin that is down can become a conduction state with spin down, and then wander away. Later, a different conduction electron with spin up can come and reside in the local orbital. This process gives the appearance of the local spin having flipped from down to up, while conduction states have flipped from up to down. Now one can see the important role of the electron–electron interaction U. The spin-flip process has two steps: (1) the departure of the old local electron with spin down, and (2) the arrival of the new local electron with spin up. If $U = 0$ these two steps are totally independent and can occcur in any order. There is no appearance of flipping, since down and up electrons come and go independently. However, once $U \gg k_B T$, then the two steps become correlated. It is energetically unfavorable to have an up and down spin electrons both on the local orbital. The two steps tend to become sequential: one leaves before the other comes. Then the local spin appears to flip.

The important parameters of the Anderson model are the bandwidths W, the conduction band density of states $g(\varepsilon)$, the local site energy ε_f, and the on-site Coulomb interaction U. There is also an energy parameter associated with the hybridization term which is defined as $\Gamma(\varepsilon)$:

$$\Gamma(\varepsilon) = \frac{\pi}{N} \sum_k V_k^2 \delta(E - \varepsilon_k) \tag{6.69}$$

$$= \pi V(\varepsilon)^2 g(\varepsilon) \tag{6.70}$$

This function is assumed to be a constant, independent of energy. Its value for electrons in f-orbitals is about 0.1 eV.

Most theories have been calculated for the case that the ground state has only one electron in the f-orbital. This corresponds to the case of cerium triply ionized (Ce^{3+}). Typical

values are $U \sim 6$ eV, $\Gamma \leq 0.1$ eV, and $\varepsilon_f \sim -0.5$ eV (Herbst and Wilkins, 1987). The zero of energy is taken to be the chemical potential. Since the spin–orbit splitting is large, one uses only the lower state which has $N_f = 6$. The energy of one local electron is ε_f, while the energy of two is $2\varepsilon_f + U$. In cerium one has that $\varepsilon_f < 0$ but $2\varepsilon_f + U > 0$. Then the cerium ion prefers to have just one electron in the local orbital. A related case is for triply ionized ytterbium which has all f-orbitals filled except one. The ground state has one f-hole. The theory is almost the same, but one uses $N_f = 8$ for the upper multiplet of the spin–orbit split f-state. When the local orbital is a d electron the values of Γ are larger.

6.2.1. No Hybridization

The term in the Hamiltonian involving U comes from the on-site Coulomb interaction between two electrons on the same atom. Usually this term is far too large to be treated by perturbation theory. It must be included in H_0, which is the noninteracting Hamiltonian. However, this term makes problems in doing the Green's function expansion. Including this term invalidates Wick's theorem for local electrons. The Green's function expansion becomes awkward. In evaluating correlation functions, each time ordering must be evaluated separately. This increases the effort, as was evident in the prior section on the Kondo effect. Here we use the noncrossing approximation (Bickers, 1987).

There are several other approaches to solving the Anderson model besides the noncrossing approximation. The method of Gunnarsson and Schönhammer (1983, 1987) was described in our second edition. Grewe and Keiter (1981) developed an alternate formalism for expanding the S-matrix for the local operators. They found formal but rigorous results for the local self-energies. Another method of finding the same results was developed by Coleman (1984) using the *slave boson* idea of Barnes (1976, 1977). One introduces a fictitious boson which is created every time there is a change in the occupancy of the local level. One can control the f occupancy by controlling the chemical potential of the slave bosons. A similar method was used earlier by Abrikosov (1965) in solving the Kondo problem. Another method is renormalization group (Hewson, 1993).

The first step is to evaluate the Hamiltonian H_0 while setting $V = 0$. This step is easy, even while including the Hubbard interaction U. The eigenvalues of the electrons, in the absence of hybridization, depend upon the number n of electrons in the f-orbital. This integer ranges in value $0 \leq n \leq N_f$. The eigenvalues are

$$E_n = n\varepsilon_f + \frac{n(n-1)}{2}U \tag{6.71}$$

where the number of pair interactions multiplying U is $n(n-1)/2$. The eigenvalue expression (6.71) is parabolic in the level occupancy n. An example is shown in Fig. 6.4. Since ε_f is usually negative, the minimum energy is found near to $n \sim \frac{1}{2} - \varepsilon_f/U$. Denote as L the value of n which has the lowest value of E_n. The f-configuration with the lowest energy has an occupancy of L electrons. Note that L is increased by changing the value of ε_f towards larger negative values, while keeping U fixed.

Mixed valence occurs when the number of f-electrons can vary between L and $L \pm 1$. In Fig. 6.4 it varies between the states with $n = 3$ and $n = 4$, which are at the same low energy. The hybridization interaction causes mixing whenever $|E_L - E_{L\pm1}|$ is a small energy. By

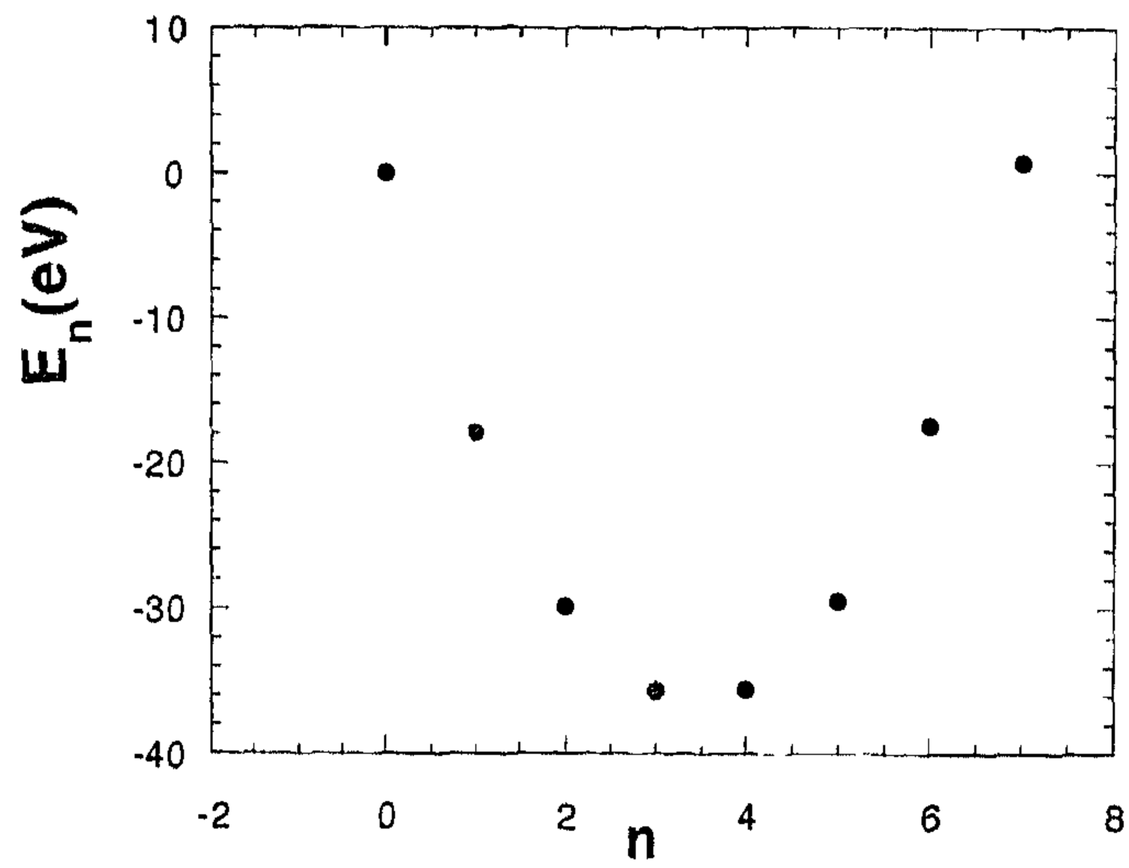


FIGURE 6.4 Eigenvalues E_n vs n for the Anderson model with no hybridization. Values assumed in the plot were: $\varepsilon_f = -18.0\,\text{eV}$ and $U = 6.0\,\text{eV}$.

"small energy" is meant values in the range of 0.1–1.0 eV. The standard nomenclature is to define the two *excitation energies* $\Delta_\pm$ out of the ground state as

$$\Delta_- = E_{L-1} - E_L = -[\varepsilon_f + (L-1)U] \tag{6.72}$$

$$\Delta_+ = E_{L+1} - E_L = \varepsilon_f + LU \tag{6.73}$$

$$U = \Delta_+ + \Delta_- \tag{6.74}$$

The definition of the Hubbard U as being equal to $\Delta_- + \Delta_+$ follows by summing the prior two equations. Equation (6.74) can also be used as a phenomenological definition. Using band structure methods to calculate Δ_- and Δ_+, one can add them to obtain a theoretical value for U.

Values of $\Delta_\pm$ have been calculated for the rare-earth metals. The results of Herbst and Wilkins (1987) are shown in Table 6.1. Generally they find one value which is small ($\Delta \approx 1.0\,\text{eV}$) while the other is larger so that $U \sim 6$–$8\,\text{eV}$. It is assumed that similar values apply to the rare-earths ions when they are in metallic compounds. Note that samarium and ytterbium have small values of Δ_+ and they have a mixed valence in many compounds. Cerium also has a mixed valence in many compounds, although its value for Δ_- does not seem to be small. An important point is that these values do not have to be on the order of thermal energies, but in fact are generally much larger: a sizable fraction of an electron volt. Even with these large values, mixed valent behavior occurs often.

Next calculate the partition function Z which is related to the thermodynamic potential $Z = \exp(-\beta\Omega)$. This quantity is needed for all of the thermodynamic averaging. Define the degeneracy factor Z_n

$$Z_n = \binom{N_f}{n} = \frac{N_f!}{n!(N_f - n)!} \tag{6.75}$$

This factor is obtained by asking for the number of different ways that n identical particles can be distributed among N_f identical states.

The ground state degeneracy of a rare-earth ion are given according to Hund's rules. They are shown in Table 6.1. The Anderson model ignores the splittings of the f-orbital due to spin–orbit and other factors, and gives a degeneracy which is far too large. The Anderson

Table 6.1 Ground staste degeneracies $(2J+1)$ for valence (3+) of rare earths according to Hund's rules: $\Delta_\pm = E_{n\pm1} - E_n$ in eV. The second entry for Eu has valence 2+. From Herbst and Wilkins (1987, used with permission)

	nf	S	L	J	$2J+1$	Δ_-	Δ_+
Ce	1	$\frac{1}{2}$	3	$\frac{5}{2}$	6	1.8	3.2
Pr	2	1	5	4	9	3.8	1.9
Nd	3	$\frac{3}{2}$	6	$\frac{9}{2}$	10	5.2	1.5
Pm	4	2	6	4	9	5.5	1.6
Sm	5	$\frac{5}{2}$	6	$\frac{5}{2}$	6	5.4	0.3
Eu	6	3	3	0	1	7.9	
Eu	7	$\frac{7}{2}$	0	$\frac{7}{2}$	8	1.9	8.6
Gd	7	$\frac{7}{2}$	0	$\frac{7}{2}$	8	8.5	3.7
Tb	8	3	3	6	13	3.2	2.2
Dy	9	$\frac{5}{2}$	5	$\frac{15}{2}$	16	4.9	1.5
Ho	10	2	6	8	17	5.8	1.4
Er	11	$\frac{3}{2}$	6	$\frac{15}{2}$	16	5.9	1.8
Tm	12	1	5	6	13	5.5	1.1
Yb	13	$\frac{1}{2}$	3	$\frac{7}{2}$	8	7.4	0.9

model is solved as given in (6.68), which has a degeneracy given by Z_n rather than Hund's rules. The partition function is

$$Z = \sum_{n=0}^{N_f} Z_n e^{-\beta E_n} \tag{6.76}$$

When $U = 0$ this summation simplifies to $Z = [1 + e^{-\beta\varepsilon_f}]^{N_f}$.

The Green's function for the f-states in the absence of the hybridization term is

$$\mathcal{G}_f(i\omega_m) = -\sum_\mu \int_0^\beta d\tau e^{-\omega_m\tau} \langle T_\tau f_\mu(\tau) f_\mu^\dagger \rangle \tag{6.77}$$

The expression for $\mathcal{G}_f$ can be written exactly for the noninteracting states. The trace is taken over the states $|n\rangle$. The notation N_f denotes the number of states for a single f-electron. The degeneracy of having n electrons in the f-orbital is Z_n as defined above. The τ integral is

$$\begin{aligned}
\mathcal{G}_f &= -\frac{1}{Z}\sum_{\mu,n} Z_n e^{-\beta E_n} \int_0^\beta d\tau e^{i\tau\omega_n} \langle n| e^{\tau H_0} f_\mu e^{-\tau H_0} |n+1\rangle \langle n+1| f_\mu^\dagger |n\rangle \\
&= -\frac{1}{Z}\sum_{\mu,n} Z_n e^{-\beta E_n} \langle n| f_\mu |n+1\rangle \langle n+1| f_\mu^\dagger |n\rangle \int_0^\beta d\tau e^{\tau(i\omega_n + E_n - E_{n+1})} \\
&= \frac{1}{Z}\sum_{\mu,n} Z_n \langle n| f_\mu |n+1\rangle \langle n+1| f^\dagger |n\rangle \frac{e^{-\beta E_n} + e^{-\beta E_{n+1}}}{i\omega_n + E_n - E_{n+1}}
\end{aligned} \tag{6.78}$$

The matrix elements $\sum_\mu \langle n| f_\mu |n+1\rangle \langle n+1| f^\dagger |n\rangle$ are easily evaluated since they are just the number of empty states into which one can add an electron into the f-orbital. If there are already n, then the number of empty spaces is $N_f - n$. In the term with $\exp(-\beta E_{n+1})$ change

summation variables $n+1 \to n$. Since $Z_n(N_f - n) = (n+1)Z_{n+1}$ these steps produce the final expression

$$\mathscr{G}_f(i\omega_n) = \frac{1}{Z}\sum_{n=0}^{N_f} Z_n e^{-\beta E_n}\left\{\frac{N_f - n}{i\omega_m + E_n - E_{n+1}} + \frac{n}{i\omega_m + E_{n-1} - E_n}\right\} \tag{6.79}$$

$$A_f(\omega) = 2\,\mathrm{Im}[G_f(\omega)] \tag{6.80}$$

$$A_f(\omega) = \frac{2\pi}{Z}\sum_{n=0}^{N_f} Z_n e^{-\beta E_n}\{(N_f - n)\delta(\omega + E_n - E_{n+1}) + n\delta(\omega + E_{n-1} - E_n)\} \tag{6.81}$$

The spectral function $A_f(\omega)$ is obtained by the analytic continuation $i\omega_m \to \omega + i\eta$ and then taking twice the imaginary part. The factors of $N_f - n$ in the first term, and n in the second, arise from the degeneracy of the state. In the state $|n\rangle$ there are n different states that can be destroyed by the lowering operator f_μ when summing over all μ. This accounts for the factor of n. Similarly, the factor $N_f - n$ is the number of empty states into which the raising operator can add an f-electron.

Let E_L be the lowest eigenvalue ($n = L$) and assume that L is neither 0 nor N_f. The most important terms in the summation are those for $n = L-1, L, L+1$. Factor $Z_L \exp(-\beta E_L)$ from each term. Energy differences $E_L - E_{L\pm1}$ are expressed as $\Delta_\pm$. These approximations give for the partition function

$$Z \approx Z_{L-1}e^{-\beta E_{L-1}} + Z_L e^{-\beta E_L} + Z_{L+1}e^{-\beta E_{L+1}} \tag{6.82}$$

$$Z \approx \mathscr{Z} Z_L e^{-\beta E_L} \tag{6.83}$$

$$\mathscr{Z} = 1 + \frac{Z_{L-1}}{Z_L}e^{-\beta(E_{L-1}-E_L)} + \frac{Z_{L+1}}{Z_L}e^{-\beta(E_{L+1}-E_L)} \tag{6.84}$$

$$\mathscr{Z} = 1 + \frac{L}{N_f + 1 - L}e^{-\beta\Delta_-} + \frac{N_f - L}{L+1}e^{-\beta\Delta_+} \tag{6.85}$$

Other terms can be ignored since they have exponential factors $\exp(-\beta\Delta E)$ with very large values of ΔE. These exponential factors are vanishing small. The same approximation for the spectral function gives six terms which have four different energy differences

$$\begin{aligned} A_f(\omega) \approx \frac{2\pi}{\mathscr{Z}}\Bigg[& L\delta(\omega + \Delta_-)(1 + e^{-\beta\Delta_-}) + (N_f - L)\delta(\omega - \Delta_+)(1 + e^{-\beta\Delta_+}) \\ & + e^{-\beta\Delta_-}\frac{L(L-1)}{N_f + 1 - L}\delta(\omega + E_{L-2} - E_{L-1}) \\ & + e^{-\beta\Delta_+}\frac{(N_f - L)(N_f - L - 1)}{L+1}\delta(\omega + E_{L+1} - E_{L+2})\Bigg] \end{aligned} \tag{6.86}$$

The approximate spectra has four peaks. Two are below the chemical potential and two above. We follow custom and ignore the last two terms in $A_f(\omega)$ and include only the peaks at $\omega = -\Delta_-, \Delta_+$. For the case that $L = 1$ then the third term is missing anyway since there is no state $L-2$. Usually it is a good approximation to neglect terms of $O(e^{-\beta\Delta_\pm})$, and in this case the approximate spectral function is

$$A_f(\omega) \approx 2\pi[L\delta(\omega + \Delta_-) + (N_f - L)\delta(\omega - \Delta_+)] \tag{6.87}$$

$$\int \frac{d\omega}{2\pi}A_f(\omega) = N_f \tag{6.88}$$

This spectral function has only two peaks: one at $\omega = \Delta_+$ and the other one at $\omega = -\Delta_-$. They have relative strength of $N_f - L$ and L, respectively. There are only two peaks, since the Green's function gives the probability for two processes: (1) adding an electron, or (2) removing an electron. In a well-defined ground state, these two processes each have a peak, and there are two peaks. Without the Hubbard energy U both peaks would occur at the same energy: there would be a single peak with weight N_f. The spectral function integrates to N_f, rather than to one, because the present definition included a summation over all of the state ν, of which there are N_f.

The solution to the Anderson model is rather easy without hybridization. The next step is to include this important process.

6.2.2. With Hybridization

An expression for the partition function is derived including the interaction V. Define the eigenstates of H_0 by $|n\rangle$ and those of $H = H_0 + V$ by $|\alpha)$.

$$H_0|n\rangle = E_n|n\rangle \tag{6.89}$$

$$H|\alpha) = \varepsilon_\alpha|\alpha) \tag{6.90}$$

The states $|n\rangle$ have n electrons in the f orbital. The nature of the states $|\alpha)$ is to be determined. In principle, they exist, and the derivation proceeds using that assumption. They form a complete set, so an insertion of $\sum_\alpha |\alpha)(\alpha|$ is equivalent to the identity operator. The derivation below proceeds by inserting this operator whenever the total Hamiltonian H appears in any expression. Using this process to evaluate the partition function gives:

$$Z = \sum_n Z_n \langle n|e^{-\beta H}|n\rangle \tag{6.91}$$

$$Z = \sum_{n\alpha} Z_n \langle n|\alpha)e^{-\beta\varepsilon_\alpha}(\alpha|n\rangle \tag{6.92}$$

$$Z = \sum_{n\alpha} Z_n |\langle n|\alpha)|^2 e^{-\beta\varepsilon_\alpha} \tag{6.93}$$

Next introduce the spectral function $A_n(\omega)$ for the many-electron state n. In prior usage the symbol $A_n(\omega)$ may have meant the spectral function of a single electron in the state n. Here the meaning is different. The symbol $|n\rangle$ denotes a state of n electrons in the f orbital, and $A_n(\omega)$ is its spectral function. It is the spectral function for a system of n electrons. It is not the spectral function of a single electron, unless $n = 1$.

$$A_n(\omega) = 2\pi\langle n|\delta(\omega - H)|n\rangle \tag{6.94}$$

$$= 2\pi \sum_\alpha |\langle n|\alpha)|^2 \delta(\omega - \varepsilon_\alpha) \tag{6.95}$$

where the complete set of states $\sum_\alpha |\alpha)(\alpha|$ is inserted in order to evaluate the delta function containing H. This formula for the spectral function can be used to evaluate our above formula for the partition function. From now on, the expression $\sum_\alpha |\langle n|\alpha)|^2$ will be evaluated as an integral over the spectral function

$$Z = \sum_n Z_n \int \frac{d\omega}{2\pi} e^{-\beta\omega} A_n(\omega) \tag{6.96}$$

The partition function can be expressed as a summation over the noninteracting states n. For each value of n, integrate over ω with the factor in the integrand of $\exp(-\beta\omega)A_n(\omega)$. The method of calculating $A_n(\omega)$ is described below.

6.2.3. Self-Energy of Electrons

The hybridization interaction V in Eq. (6.68) describes the processes whereby a conduction electron can hop on or off the f shell of the local atom. This hopping preserves the spin and orbital quantum numbers of the electron. At first it appears to have nothing to do with spin fluctuations. As explained in the introduction to this section, the fluctuations in occupancy can change the spin of the local electron. The local spin fluctuations are due to the hopping process.

This process leads to several important new effects. The first is a new energy parameter defined as Γ in Eq. (6.69). It is taken to be a constant. The second effect is that the resonance lines at $\omega = -\Delta_-, \Delta_+$ become broadened by an amount proportional to Γ. The third new effect is that the Kondo resonance peak appears in $A_f(\omega)$ near to zero energy. It is an important feature in the Anderson model, which is due to spin fluctuations.

The prior subsection discussed the energy levels associated with having n electrons in the f state. Introduce a Green's function for this state

$$G_n(\omega) = \frac{1}{\omega - E_n - \Sigma_n(\omega)} \tag{6.97}$$

Most of the Green's functions in this book are for a single electron. There are also two-particle correlation functions, such as density–density or current–current correlation functions. The present Green's function is neither of these (unless $n = 1$). It is a Green's function for a state of n electrons in the f orbital, where values of n have the range $0 \le n \le N_f$. Similarly, assume there is a self-energy Σ_n for this n-particle Green's function, which can be evaluated using Dyson's equation.

The hybridization interaction causes the self-energy $\Sigma_n(\omega)$. This function is calculated using the noncrossing approximation (Bickers, *et al.*, 1987). The result will be derived by an intuitive argument. A rigorous definition is provided below. First consider the result for the self-energy in the second order of perturbation theory

$$\Sigma_n(\omega) = nS'^{(0)}_{n-1}(\omega) + (N_f - n)S^{(0)}_{n+1}(\omega) \tag{6.98}$$

$$S^{(0)}_{n+1}(\omega) = \frac{1}{N}\sum_k \frac{V_k^2 n_k}{\omega - (E_{n+1} - \varepsilon_k) + i\eta} \tag{6.99}$$

$$S'^{(0)}_{n-1}(\omega) - \frac{1}{N}\sum_k \frac{V_k^2(1 - n_k)}{\omega - (E_{n-1} + \varepsilon_k) + i\eta} \tag{6.100}$$

There are two terms. The first term $(f^\dagger C_k)$ in V adds an electron to the f levels, which changes n to $n + 1$, while destroying a band electron in the state k. This process is proportional to the occupation number n_k of the band electrons, and gives the term $S^{(0)}_{n+1}(\omega)$. Again the factor of $N_f - n$ is the number of empty states into which the electron can be inserted into the f orbital with n electrons. The second term $(C_k^\dagger f)$ in V takes an electron away from the f level, which changes n to $n - 1$, while adding an electron to the band. This process is proportional to the probability $(1 - n_k)$ that the band state is unoccupied. It gives the term $S'^{(0)}_{n-1}(\omega)$. These two processes provide the two terms in the above expression for Σ_n.

Note that the arguments of $S^{(0)}$ are Green's functions G_n without the self-energies.

$$S_n^{(0)}(\omega) = \frac{1}{N}\sum_k V_k^2 n_k G_n^{(0)}(\omega + \varepsilon_k) \tag{6.101}$$

The "non-crossing approximation" (NCA) is to evaluate these Green's functions while including these self-energies. Put the self-energies in the denominators

$$\Sigma_n(\omega) = nS'_{n-1}(\omega) + (N_f - n)S_{n+1}(\omega) \tag{6.102}$$

$$S_{n+1}(\omega) = \frac{1}{N}\sum_k V_k^2 n_k G_{n+1}(\omega + \varepsilon_k) \tag{6.103}$$

$$S'_{n-1}(\omega) = \frac{1}{N}\sum_k V_k^2 (1 - n_k) G_{n-1}(\omega - \varepsilon_k) \tag{6.104}$$

The above nonlinear equations comprise a self-consistent definition for the self-energies Σ_n. The values of n range over a limited number of possibilities ($0 \le n \le N_f$), so only a few functions of energy need to be determined. They can be solved by iteration on the computer. The temperature enters into the expressions through the Fermi–Dirac occupation numbers n_k. The Kondo resonance is quite temperature dependent, and must be calculated anew for each temperature.

The results are illustrated by a simple example. Assume a density of states with a Lorentzian shape.

$$g(\varepsilon) = \frac{2D}{\varepsilon^2 + D^2} \tag{6.105}$$

and place the chemical potential at $\varepsilon = 0$. This choice has the advantage of not introducing band edges. The parameter D is an effective band width and $\Gamma = \pi g(0)V(0)^2 = 2\pi V(0)^2/D$. In $S'_n(\omega)$, change variables $\varepsilon \to -\varepsilon$ which now makes $S'_n = S_n$. The main integral to evaluate is

$$S_n(\omega) = \frac{\Gamma}{\pi}\int \frac{d\varepsilon}{1 + (\varepsilon/D)^2} n_F(\varepsilon) G_n(\omega + \varepsilon) \tag{6.106}$$

Next discuss the derivation of Eq. (6.102). Again define $|n\rangle$ as the exact eigenstate of H_0 while $|\alpha)$ is the exact eigenstate of $H = H_0 + V$. The definition of the Green's function is

$$G_n(\omega) = \langle n| \frac{1}{\omega - H} |n\rangle \tag{6.107}$$

$$= \sum_\alpha \frac{|(\alpha|n\rangle|^2}{\omega - \varepsilon_\alpha + i\eta} \tag{6.108}$$

Again a summation over the states $|\alpha)(\alpha|$ was inserted in order to evaluate the energy denominator with the Hamiltonian H. This result could also be obtained using the Lehmann representation.

The self-energy Σ_n can be obtained by a similar argument. One starts with the state $|n\rangle$ and evaluates the self-energy in the second order of perturbation theory. The interaction V is regarded as small, so an expression valid to $O(V^2)$ is considered adequate. The general expression is

$$\Sigma_n(\omega) = \sum_I \frac{|(I|V|n\rangle|^2}{\omega - \varepsilon_I} \tag{6.109}$$

The intermediate states $|I)$ are taken to be the set $|\alpha)$. The perturbation V has two terms: one $(f^\dagger C_k)$ adds an electron to the f orbital while removing an electron from the conduction band. The other term $(C_k^\dagger f)$ removes an electron from the f orbital while adding one to the conduction band. These two processes generate the two terms:

$$\Sigma_n(\omega) = \frac{1}{N}\sum_{k\alpha} V_k^2 \left\{ \frac{n_k|(\alpha|f^\dagger|n\rangle|^2}{\omega - (\varepsilon_\alpha - \varepsilon_k)} + \frac{(1-n_k)|(\alpha|f|n\rangle|^2}{\omega - (\varepsilon_\alpha + \varepsilon_k)} \right\} \tag{6.110}$$

$$= \frac{1}{N}\sum_{k\alpha} V_k^2 \left\{ \frac{n_k|(\alpha|n+1\rangle\langle n+1|f^\dagger|n\rangle|^2}{\omega - (\varepsilon_\alpha - \varepsilon_k)} + \frac{(1-n_k)|(\alpha|n-1\rangle\langle n-1|f|n\rangle|^2}{\omega - (\varepsilon_\alpha \varepsilon_k)} \right\} \tag{6.111}$$

$$= \frac{1}{N}\sum_k V_k^2 \{(N_f - n)n_k G_{n+1}(\omega + \varepsilon_k) + n(1-n_k)G_{n-1}(\omega - \varepsilon_k)\}$$

The summation over intermediate states $(I|$ includes the exact states $(\alpha|$ and the eigenstates of the band electrons. They are summed explicitly in the second line. A factor of $|n \pm 1\rangle\langle n \pm 1|$ was also inserted, to use the fact that $\langle n+1|f^\dagger|n\rangle$ and $\langle n-1|f|n\rangle$ are well defined. This produces in (6.111) an expression which is obviously the Green's function. The last equation, in the above set, is the expression for the self-energy in the non-crossing approximation.

A similar technique can be used to derive the spectral function $A_f(\omega)$. Start the derivation with the general definition of the one-electron Green's function:

$$\mathscr{G}_f(ik_n) = -\frac{1}{Z}\sum_n Z_n \int_0^\beta d\tau e^{ik_n\tau}\langle n|e^{-(\beta-\tau)H} f e^{-\tau H} f^\dagger|n\rangle \tag{6.112}$$

This Green's function is the usual one, similar to those in other chapters, where the definition entails adding or removing a single electron from the system. It is the function which is of physical interest, since it directly relates to the density of states and other quantities. On the other hand, the Green's function $G_n(ik_n)$ is that of the n-particle state. It is the result of adding n electrons to the system. It generally is hard to measure. In the present case G_n is useful since it is used to calculate $\mathscr{G}_f$.

The expression in brackets can be written exactly by inserting summations over exact states $|\alpha)$ and $|\lambda)$, as well as noninteracting states $|n+1\rangle$ and $|n'+1\rangle$.

$$\sum_{\alpha\lambda n'} \langle n|\alpha)e^{-(\beta-\tau)\varepsilon_\alpha}(\alpha|n'\rangle\langle n'|f|n'+1\rangle\langle n'+1|\lambda)e^{-\tau\varepsilon_\lambda}(\lambda|n+1\rangle\langle n+1|f^\dagger|n\rangle \tag{6.113}$$

The usual approximation is to restrict the summation to the terms with $n' = n$. Gunnarsson and Schönhammer (1983) consider the terms with $n' = n \pm 1$ and find they make a small contribution. They are a type of vertex correction. Setting $n' = n$ simplifies the expression. After performing the integral over τ

$$\mathscr{G}_f(ik_n) = \frac{1}{Z}\sum_n Z_n(N_f - n)\sum_{\alpha\lambda} |(\alpha|n\rangle|^2|(\lambda|n+1\rangle|^2 \frac{e^{-\beta\varepsilon_\alpha} + e^{-\beta\varepsilon_\lambda}}{ik_n + \varepsilon_\alpha - \varepsilon_\lambda}$$

The matrix elements such as $|(\alpha|n\rangle|^2$ and $|(\lambda|n+1\rangle|^2$ are expressed in terms of the spectral functions A_n and A_{n+1} which produces the result

$$\mathscr{G}_f(ik_n) = \frac{1}{Z}\sum_n Z_n(N_f - n)\int \frac{d\varepsilon}{2\pi} A_n(\varepsilon)\int \frac{d\varepsilon'}{2\pi} A_{n+1}(\varepsilon')\frac{e^{-\beta\varepsilon} + e^{-\beta\varepsilon'}}{ik_n + \varepsilon - \varepsilon'}$$

Taking the imaginary part of the retarded function gives the delta function $\delta(\omega + \varepsilon - \varepsilon')$ which can be used to eliminate one of the energy integrals. After rearranging a bit,

$$A_f(\omega) = \frac{1}{Z}\sum_{n=0}^{N_f} Z_n \int \frac{dE}{2\pi} e^{-\beta E} A_n(E)\{nA_{n-1}(E-\omega) + (N_f - n)A_{n+1}(E+\omega)\} \tag{6.114}$$

The final equation is the result (6.114) for the spectral function $A_f(\omega)$. Note that the integral for $A_f(\omega)$ is a convolution of the spectral functions for adjacent values of n. In practice, the only important terms in the summation are those with $n = L-1, L, L+1$. Again ignoring terms of $O(e^{-\beta\Delta_\pm})$ then this expression simplifies to

$$\begin{aligned} A_f(\omega) &\approx \frac{1}{\tilde{Z}_L}\int \frac{dE}{2\pi} e^{-\beta E} A_L(E)\{LA_{L-1}(E-\omega) + (N_f - L)A_{L+1}(E+\omega)\} \\ \tilde{Z}_L &\approx \int \frac{dE}{2\pi} e^{-\beta E} A_L(E) \end{aligned} \tag{6.115}$$

A numerical example of these equations is presented. For cerium, one typically chooses values such as $\varepsilon_f = -0.5$ eV, $U = 6.0$ eV, and $\Gamma = 0.1$ eV. This gives $E_0 = 0, E_1 = -0.5$ eV, $E_2 = 5.0$ eV, and the other levels are unimportant. The width parameter in the Lorentzian density of states is $D = 1.0$ eV. The energy differences are $\Delta_- = 0.5$ eV and $\Delta_+ = 5.5$ eV. The ground state has $L = 1$ and the mixed valence is between $L = 1$ and $L = 0$. Figure 6.5 shows these spectral functions evaluated at $T = 300$ K. The spectral function for the ground state $A_1(E)$ is almost a pure Lorentzian with small width. This small width is provided by the small value of Im Σ_1 due to its interactions with the neigboring state with $L \pm 1$. From Eq. (6.102)

$$-2\,\mathrm{Im}[\Sigma_1(\omega)] = \frac{1}{N}\sum_k V_k^2[(1-n_k)A_0(\omega - \varepsilon_k) \tag{6.116}$$

$$+ (N_f - 1)n_k A_2(\omega + \varepsilon)] \tag{6.117}$$

$$= \Gamma \int \frac{d\varepsilon}{1 + (\varepsilon/D)^2} n_F(\varepsilon)[A_0(\omega+\varepsilon) + (N_f - 1)A_2(\omega + \varepsilon)]$$

The integration variable was changed in the first term $\varepsilon \to -\varepsilon$ which is only possible when $\mu = 0$ and with a symmetric density of states. Both terms in the integral contribute about equally: Although E_2 is further away in energy, and less effective, the factor of $N_f - 1$

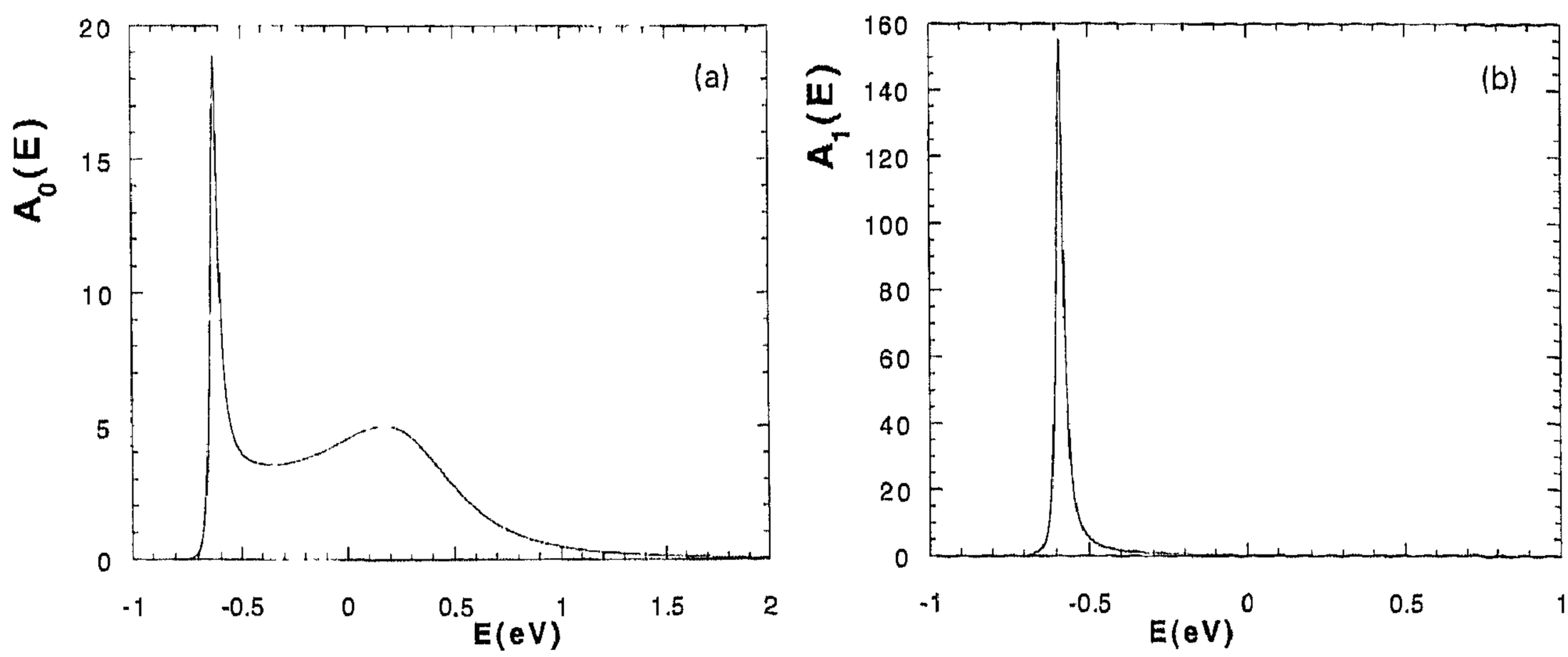


FIGURE 6.5 Spectral functions for cerium at $T = 300$ K: (a) A_0, (b) A_1. Energy units are eV.

multipling A_2 increases the importance of this term. The $1/N_f$ expansion is based upon the premise that terms of $O(N_f)$ are much larger than those of $O(1)$ and the latter can be neglected. Our experience is that this assumption is not accurate numerically.

The spectral function for A_0 has a main peak at E_0 which is nearly Lorentzian. The broadening is caused by Γ. However, there is also a narrow peak at the energy E_1 of the ground state. It is due to the self energy $\mathrm{Im}[\Sigma_0]$ defined by A_1: the latter function has a peak there. The peak at the ground state energy, in A_0, is evidence that this state is fluctuating with the ground state. The spectral function of A_2 does not show a peak at the energy E_1, since E_2 is too far away.

The spectral function $A_f(\omega)$ is found by the integral in Eq. (6.114). It is shown in Fig. 6.6. Since both A_1 and A_0 have sharp peaks at the same energy E_1, their convolution has a sharp peak near to zero energy. The peak near to zero energy is the Kondo resonance. The Kondo resonance has a narrow width. This narrow width is close in value to $k_B T_K$, where T_K is the Kondo temperature. The Kondo phenomena is caused by this sharp peak, and the width of the peak—its characteristic energy—defines the energy scale which in turn defines the temperature scale. The Kondo temperature for this case is about $T_K = 300\,\mathrm{K}$. The height of the Kondo resonance decreases rapidly with increasing temperature.

The Kondo resonance was first derived in the prior section, where it arose from the spin-flip scattering of a conduction electron from a local impurity spin. The present Kondo resonance is the same peak, calculated from the Anderson model rather than from the Kondo model. In both cases, it arises from the interaction between the conduction and local electrons. In the Anderson model, the Hubbard U appears not to play a major role. It only enters into the energy denominators. However, this term is the major reason for the Kondo resonance. Setting $U = 0$ causes the energy denominators to become equal. The Fermi factors cancel in the numerator of the self-energy expression, and one has a simple function with dull features. For $U = 0$ the Kondo resonance is not present. The energy U for pair interactions is the major cause of the Kondo resonance. The Kondo resonance is the major macroscopic manifestation of the spin-flip scattering. This resonance explains the unusual behavior in the heat capacity and other measurements.

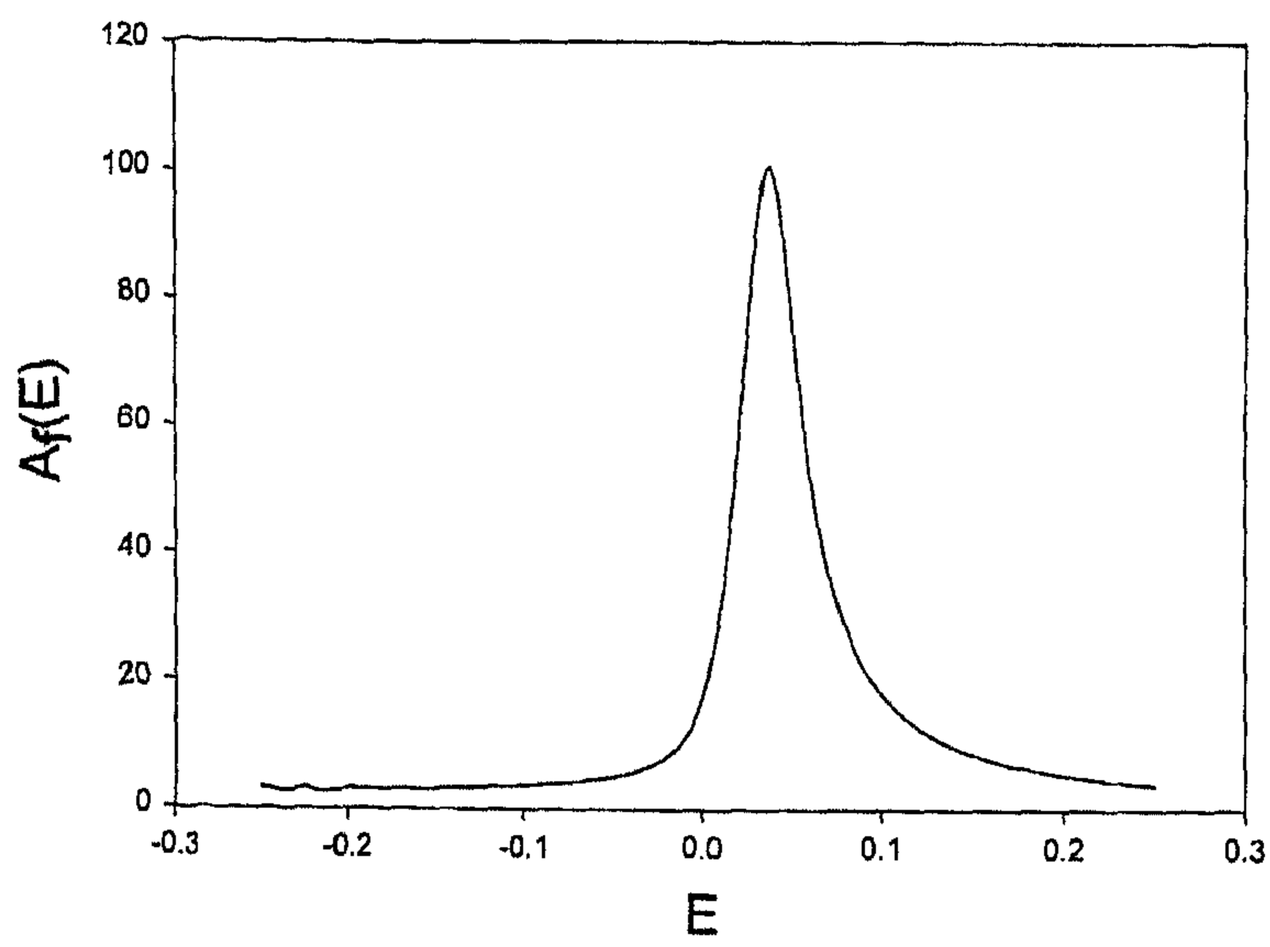


FIGURE 6.6 Kondo resonance as a function of energy for cerium. Note the expanded energy scale.

As an example, calculate the heat capacity for a system with a concentration c of such Kondo resonances. This calculation uses the formula

$$C(T) = \int dE A_f(E) E \left(-\frac{dn_F(E)}{dE} \right) \tag{6.118}$$

The heat capacity rises linearly with temperature at small temperatures. The formula for the free-electron gas $C \sim m^* k_B^2 T / E_F$ is used to estimate the effective mass m^*. One finds a value for this system of $m^* = x m_e$, where m_e is the mass of the electron and x is very large. The system appears to have a very large effective mass. This result is the origin of the name "heavy-fermion". The fermions are not that heavy. There is just a peak in the density of states due to the spin fluctuations, which explains the large value of the heat capacity at low temperature.

The low-temperature properties were not calculated correctly. If there is a concentration c of such cerium atoms, then at low temperature the system develops a band of f-states, as the f-electron hops coherently from cerium to cerium. The f-electrons develop their own Fermi surface. This subject is rather interesting, and is reviewed by Zwicknagl (1992). The band of f-states vanishes at higher temperature, as hopping becomes less probable. This behavior is an example of electronic polaron model of Holstein (see Liu, 1988).

These results can be used to calculate the electrical resistivity $\rho(T)$ and the Seebeck coefficient $S(T)$. The usual assumption is that the scattering rate for the conduction band electrons—their inverse lifetime—is proportional to a constant term from impurity scattering, a term from phonon scattering, and a term from scattering from the local f states. The first two terms are lumped into $\tau_0(T)$. The latter term is taken to be proportional to their concentration c and the spectral function A_f along with another constant u

$$\frac{1}{\tau(\omega)} = \frac{1}{\tau_0(T)} + cuA_f(\omega) \tag{6.119}$$

$$\tau(\omega) = \frac{\tau_0(T)}{1 + cu\tau_0 A_f(\omega)} \tag{6.120}$$

$$\sigma(T) = e\sigma_1 \int dE \tau(E) g(E) \left(-\frac{dn_F(E)}{dE} \right) \tag{6.121}$$

$$T\sigma S = \sigma_1 \int dE E \tau(E) g(E) \left(-\frac{dn_F(E)}{dE} \right) \tag{6.122}$$

where σ_1 is a constant. The factor of $g(E)$ is the density of states. The factor of the square of the Fermi velocity is part of σ_1. This expression was derived by Mahan (1997).

The Kondo peak has an energy scale very different to that of the scattering due to phonons or impurities, and the latter contributions are treated as constants. Note that the effect of the Kondo resonance is to cut a hole in the energy spectrum of $\tau(\omega)$ since the factor of $A_f(\omega)$ is found in the denominator. Figure 6.7 shows the calculated results of $S(T)$ for this model for cerium. Calculations of this kind were first done by Bickers *et al.* (1987). They show there is nearly a universal curve for the different cases when plotting the results on an axis of the logarithm of T/T_K. Different choices of the model parameters (ε_f, U, Γ, N_f) give different values for the Kondo temperature T_K. The curve for the Seebeck has a universal shape and maximum value ($\sim$100 μV/K) when plotted on a log scale. Note that S is the ratio of two integrals: If the concentration c of f levels is large, then the factor of c cancels between the two integrals. This cancellation is the reasons the curve for S is universal.

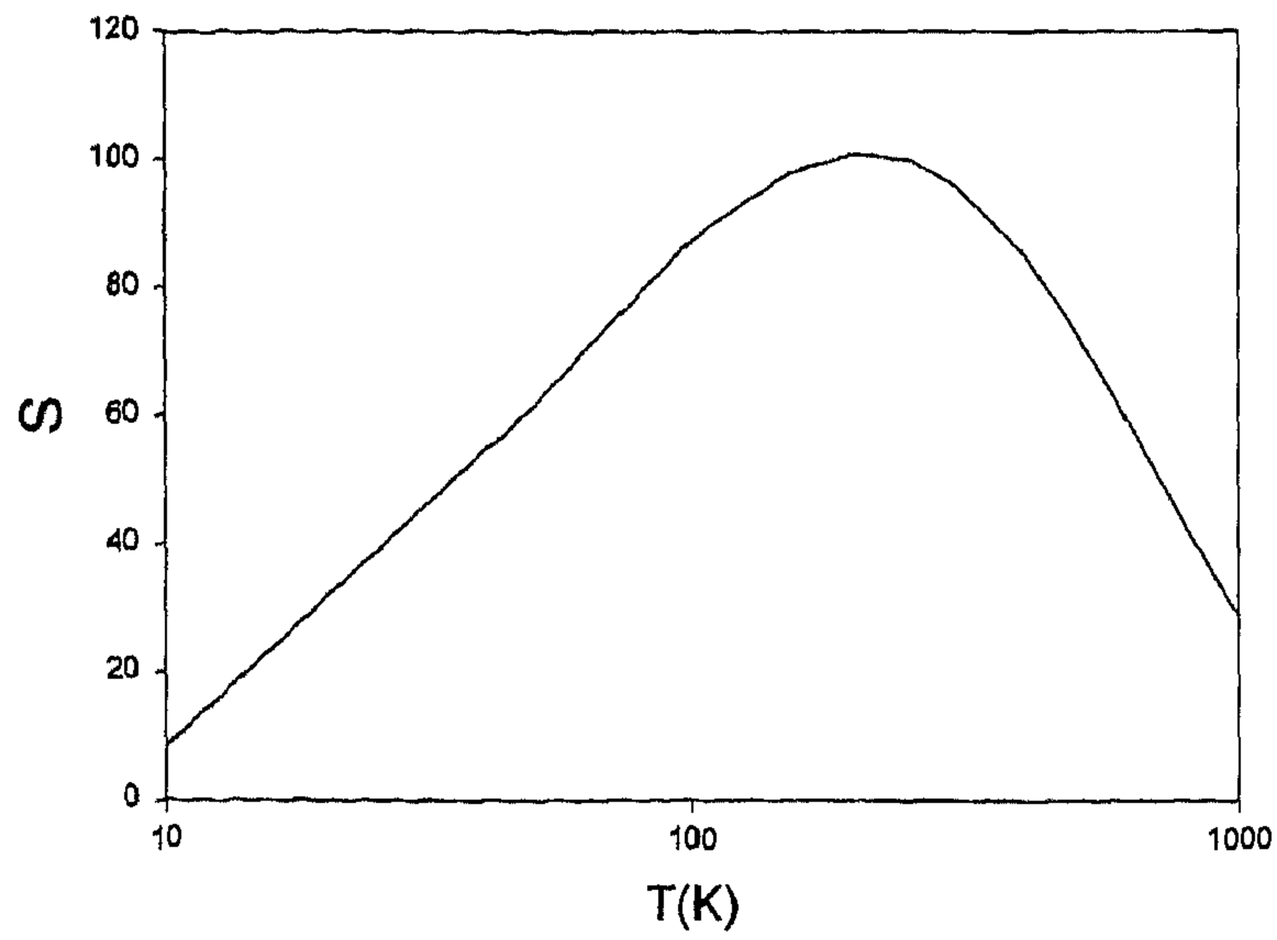


FIGURE 6.7 Seebeck coefficient $S(T)$ for a model appropriate to cerium.

Figure 6.8 shows the Seebeck coefficient for $CePd_3$ and $YbAl_3$ plotted as a function of temperature using data from Jaccard and Sierro (1982). Both the peak values, and the general shape agree with the model calculations. These Seebeck coefficients are large for metals. Most metals have values in the range of 2–5 μV/K. The Seebeck coefficient is large whenever $g(E)\tau(E)$ has sharp, asymmetric, structure within $2k_BT$ of the chemical potential (Mahan and Sofo, 1996). In this case, the sharp structure is the Kondo resonance.

The Anderson model describes spin fluctuations. There are also charge fluctuations as the electron hops on and off of the local orbital. The present theory neglects them. They could be described by including a term in the Anderson Hamiltonian of the form

$$V_{cf} = \sum_{\nu} f_{\nu}^{\dagger} f_{\nu} \sum_{\mathbf{k}\mathbf{k}'s} V_{kk'} C_{\mathbf{k}s}^{\dagger} C_{\mathbf{k}'s} \tag{6.123}$$

There is a Coulomb interaction between the local f electrons and band electrons. The latter scatter from $\mathbf{k}'$ to $\mathbf{k}$. The matrix element, and the scattering, depends upon how many local electrons are present. One can call the ion valence Z and the sum over local electrons controls the value of Z. When an electrons hops, say, off the orbital, then this sum changes. The theory of this charge fluctuation encounters singularities identical to the Anderson renormalization in

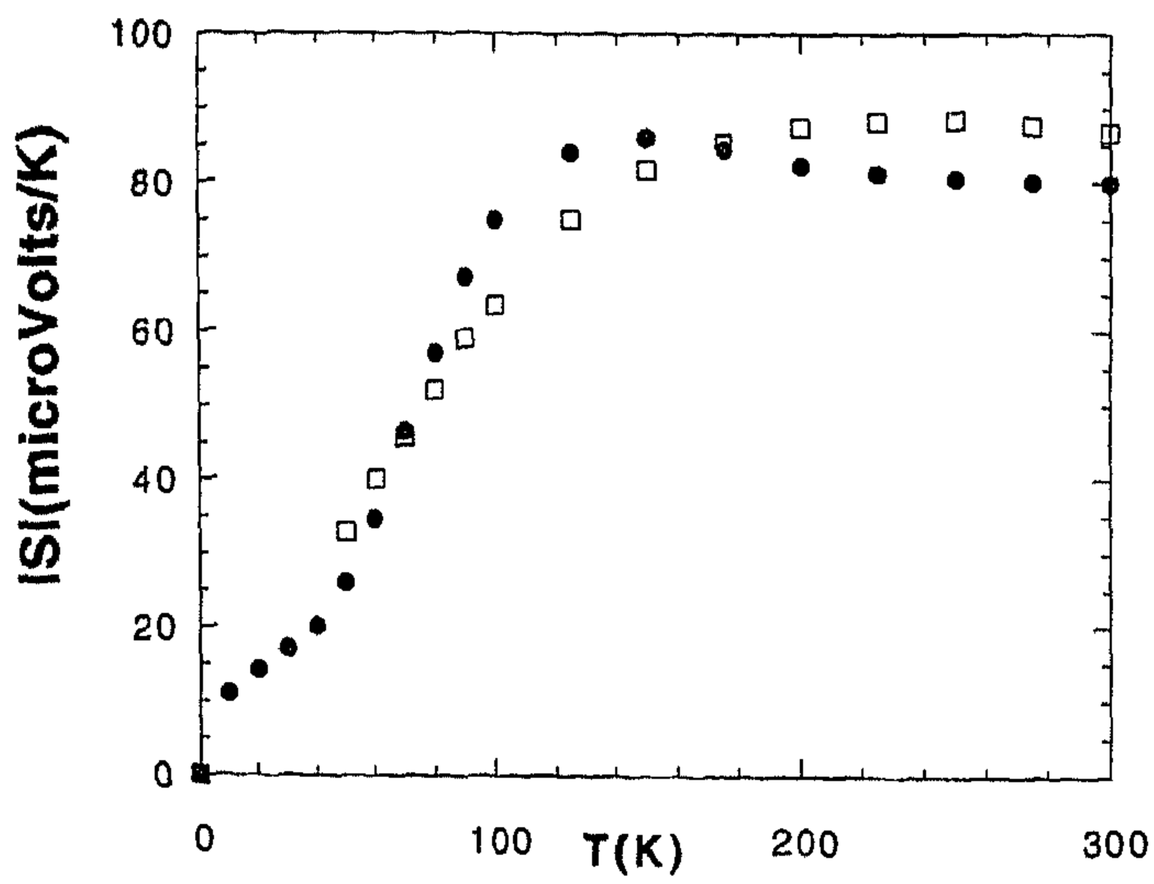


FIGURE 6.8 Absolute Seebeck coefficient as a function of temperature for $CePd_3$ and $YbAl_3$. The results for cerium are positive and for ytterbium are negative.

the x-ray edge problem, which is discussed in Chapter 9. They occur whenever one adds or removes a local electron which has a Coulomb interaction with the local electrons. This theory is discussed by Liu (1988).

6.3. HUBBARD MODEL

The Hubbard model has become the standard Hamiltonian for the investigation of electron behavior in metals where the correlation effects are quite strong. Many analytical and numerical techniques have been applied to this Hamiltonian. It is routine for young theorists to announce that they have "solved the Hubbard model". In spite of this huge effort, extending over many years, there are many features of this Hamiltonian in two and three dimensions which are not understood.

This section discusses the Hubbard Hamiltonian H_H and also the extended Hubbard Hamiltonian H_{xH}

$$H_H = -t \sum_{j\delta\sigma} C^\dagger_{j+\delta,\sigma} C_{j\sigma} + U \sum_j n_{j\uparrow} n_{j\downarrow} \tag{6.124}$$

$$n_{j\sigma} = C^\dagger_{j\sigma} C_{j\sigma} \tag{6.125}$$

$$H_{xH} = H_H + \tfrac{1}{2} \sum_{j\neq l,\sigma\sigma'} n_{j\sigma} n_{l\sigma'} V(\mathbf{R}_j - \mathbf{R}_l) \tag{6.126}$$

$$V(R) \approx \frac{e^2}{R} \tag{6.127}$$

The summation j is over lattice sites. The summation δ is over the neighbors of j. The summations over σ, σ' are over spin indices. The term t is a hopping overlap energy between first neighbors. It is assumed to be the same for all neighbors. This choice implies s-wave symmetry: the orbitals are the same in all directions. Each site has the same symmetry. It is possible to take a model with more complications, but the present one is hard enough.

The extended Hubbard Hamiltonian includes the Coulomb interaction between electrons at distance sites. This term is essential to obtain the correct charge fluctuations. The pure Hubbard Hamiltonian H_H does a good job describing spin fluctuations, but does a poor job on charge fluctuations.

The Hubbard model has several energy scales. The first is the on-site Coulomb interaction U, and the second is the bandwidth $W = 2zt$, where z is the number of nearest neighbors: it is the number of terms in the summation over δ. The behavior is very different depending upon the ratio W/U. If this ratio is large, then the value of U is small compared to the bandwidth. Here the free-electron model of metals works well. The case of simple metals such as aluminum or sodium have $W \gg U$. The other limit is when $U \gg W$. Here electron correlations are dominent. This case is only poorly understood. It is known that if the band is exactly half-full, with one electron per site, then the system will be antiferromagnetic when $U \gg W$. There have been many attempts, beginning with Penn (1966), to derive a phase diagram at zero temperature to predict all of the different order parameters at different concentrations n and different values of U/W. Phase diagrams are discussed in the next section.

6.3.1. Spin and Charge Separation

Both the Hubbard model, and the extended Hubbard model, can be written in a form which demonstrates the separation of charge and spin degrees of freedom. The extended Hubbard model is discussed here. The Hubbard model is obtain, at any point in the derivation, by merely setting to zero the interaction $V(R)$.

The first step is to change the Hamiltonian to wave vector space.

$$C_{j\sigma} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_j} C_{\mathbf{k}\sigma} \tag{6.128}$$

$$-t \sum_{j\delta\sigma} C^{\dagger}_{j+\delta,\sigma} C_{j\sigma} = \sum_{\mathbf{k}\sigma} \varepsilon(\mathbf{k}) C^{\dagger}_{\mathbf{k}\sigma} C_{\mathbf{k}\sigma} \tag{6.129}$$

$$\varepsilon(\mathbf{k}) = -t \sum_{\delta} e^{i\mathbf{k}\cdot\delta} \tag{6.130}$$

The term in $\varepsilon(\mathbf{k})$ is the kinetic energy. The other terms are the various Coulomb interactions. The Fourier transform of the extended Coulomb interaction has an interesting form

$$v(\mathbf{q}) = \sum_{j\neq 0} e^{i\mathbf{q}\cdot\mathbf{R}_j} V(\mathbf{R}_j) \tag{6.131}$$

$$0 = \sum_{\mathbf{q}} v(\mathbf{q}) \tag{6.132}$$

The term with $j \neq 0$ is omitted from the summation since it is the Hubbard interaction U. The term U is included separately from $v(\mathbf{q})$. The second formula, in the above set, shows that the summation of $v(\mathbf{q})$ over all values of $\mathbf{q}$ gives zero. This happens because the summation of $v(\mathbf{q})$ over all $\mathbf{q}$ gives $V(R = 0)$ and that term is omitted from the definition of $v(\mathbf{q})$.

The actual summation must be done with Ewald methods. At small values of q, $v(\mathbf{q}) \to 4\pi e^2/(q^2 V_0)$ where V_0 is the volume of the unit cell. The potential $v(\mathbf{q})$ behaves as a normal Coulomb interaction at small values of q. However, since the summation of $v(\mathbf{q})$ over the crystal Brillouin zone of $\mathbf{q}$ values must give zero, then every region of positive value must be cancelled by another region where $v(\mathbf{q})$ is negative. Generally the negative regions of $v(\mathbf{q})$ occur at the points in the Brillouin zone with larger values of $\mathbf{q}$. Negative values occur from the phase factor $\exp(i\mathbf{q}\cdot\mathbf{R}_j)$ in Eq. (6.131). It can be negative when the exponent is near to π radians.

The Coulomb interactions are written in terms of the density operators of the electrons. It is convenient to introduce the electron density operator $\rho_\sigma(\mathbf{q})$ of a single spin, where $\sigma = \uparrow, \downarrow$. The charge density operator $\rho_c(\mathbf{q})$ is the sum of both spin components, while the spin density operator $\rho_s(\mathbf{q})$ is their difference

$$\rho_\sigma(\mathbf{q}) = \sum_{\mathbf{k}} C^{\dagger}_{\mathbf{k}+\mathbf{q},\sigma} C_{\mathbf{k}\sigma} \tag{6.133}$$

$$\rho_c(\mathbf{q}) = \rho_\uparrow(\mathbf{q}) + \rho_\downarrow(\mathbf{q}) \tag{6.134}$$

$$\rho_s(\mathbf{q} = \rho_\uparrow(\mathbf{q}) - \rho_\downarrow(\mathbf{q}) \tag{6.135}$$

The terms in the Hubbard interaction and in the extended Coulomb interaction can be written in terms of these operators.

$$U\sum_j nj_\uparrow n_{j\downarrow} = \frac{U}{N}\sum_{\mathbf{q}} \rho_\uparrow(\mathbf{q})\rho_\downarrow(-\mathbf{q}) \tag{6.136}$$

$$= \frac{U}{4N}\sum_{\mathbf{q}}[\rho_c(\mathbf{q})\rho_c(-\mathbf{q}) - \rho_s(\mathbf{q})\rho_s(-\mathbf{q})] \tag{6.137}$$

$$\tfrac{1}{2}\sum_{j\neq l,\sigma\sigma'} n_{j\sigma}n_{l\sigma'}V(R_{jl}) = \frac{1}{2N}\sum_{\mathbf{q}} v(\mathbf{q})\rho_c(\mathbf{q})\rho_c(-\mathbf{q}) \tag{6.138}$$

The two interactions can be combined into a charge and a spin interaction

$$V_c = \frac{1}{2N}\sum_{\mathbf{q}}[v(\mathbf{q}) + \tfrac{1}{2}U]\rho_c(\mathbf{q})\rho_c(-\mathbf{q}) \tag{6.139}$$

$$V_s = -\frac{U}{4N}\sum_{\mathbf{q}} \rho_s(\mathbf{q})\rho_s(-\mathbf{q}) \tag{6.140}$$

It is interesting that the effective interaction for charge fluctuations is $v_c(\mathbf{q}) = v(\mathbf{q}) + U/2$. This quantity can also be negative at large values of $\mathbf{q}$. There is a simple proof due to Chen and Mattis (1991) that $v(\mathbf{q}) + U$ must always be positive. Assume that the charge distribution on an atomic site is $\phi(r)$, then

$$v_T(\mathbf{q}) = U + v(\mathbf{q}) \tag{6.141}$$

$$v_T(\mathbf{q}) = \int d^3r d^3r' \phi(r)\phi(r')\sum_j \frac{e^2}{|\mathbf{r} - \mathbf{r}' + \mathbf{R}_j|} e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{6.142}$$

The summation over j includes $\mathbf{R}_j = 0$ since that is U. The summation can be written as a periodic potential and therefore expanded in reciprocal lattice vectors $\mathbf{G}$

$$\sum_j \frac{e^2}{|\mathbf{r} - \mathbf{r}' + \mathbf{R}_j|} e^{-\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}'+\mathbf{R}_j)} = \sum_{\mathbf{G}} \frac{4\pi e^2}{V_0|\mathbf{q} + \mathbf{G}|^2} e^{-i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}')} \tag{6.143}$$

$$v_T(\mathbf{q}) = \sum_{\mathbf{G}} \frac{4\pi e^2}{V_0|\mathbf{q} + \mathbf{G}|^2} |\tilde{\phi}(\mathbf{q} + \mathbf{G})|^2 \tag{6.144}$$

$$\tilde{\phi}(\mathbf{q}) = \int d^3r e^{i\mathbf{q}\cdot\mathbf{r}}\phi(r)$$

The right-hand side of Eq. (6.144) is a sum of positive terms, so it is strictly positive. Although $v(\mathbf{q}) + U$ is strictly positive, the effective Coulomb interaction $v_c(\mathbf{q}) = v_T - U/2 = v(\mathbf{q}) + U/2$ may be negative at large $\mathbf{q}$.

The next step is to evaluate the correlation function of these operators. Assume the system is nonmagnetic, so the average spin-up and spin-down correlations are the same. First assume $V_c = V_s = 0$ then

$$P_c(\mathbf{q}, iq_n) = \frac{1}{\nu}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau \rho_c(\mathbf{q}, \tau)\rho_c^\dagger(\mathbf{q}, 0)\rangle \tag{6.146}$$

$$= -\frac{1}{\nu}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau[\rho_\downarrow(\mathbf{q}, \tau)\rho_\downarrow^\dagger(\mathbf{q}, 0) + \rho_\uparrow(\mathbf{q}, \tau)\rho_\uparrow^\dagger(\mathbf{q}, 0)]\rangle \tag{6.147}$$

$$= P^{(1)}(\mathbf{q}, iq_n) \tag{6.148}$$

$$P_s(\mathbf{q}, iq_n) = -\frac{1}{\nu}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau \rho_s(\mathbf{q}, \tau)\rho_s^\dagger(\mathbf{q}, 0)\rangle \tag{6.149}$$

$$= -\frac{1}{\nu}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau[\rho_\uparrow(\mathbf{q}, \tau)\rho_\uparrow^\dagger(\mathbf{q}, 0) + \rho_\downarrow(\mathbf{q}, \tau)\rho_\downarrow^\dagger(\mathbf{q}, 0)]\rangle \tag{6.150}$$

$$= P^{(1)}(\mathbf{q}, iq_n) \tag{6.151}$$

$$P_{cs}(\mathbf{q}, iq_n) = -\frac{1}{\nu}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau \rho_c(\mathbf{q}, \tau)\rho_s^\dagger(\mathbf{q}, 0)\rangle \tag{6.152}$$

$$= -\frac{1}{\nu}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau[\rho_\uparrow(\mathbf{q}, \tau)\rho_\uparrow^\dagger(\mathbf{q}, 0) - \rho_\downarrow(\mathbf{q}, \tau)\rho_\downarrow^\dagger(\mathbf{q}, 0)]\rangle \tag{6.153}$$

$$= 0 \tag{6.154}$$

where $P^{(1)}(\mathbf{q}, iq_n)$ is the usual polarization function for the electron gas, which was discussed extensively in Chapter 5. Terms such as $\langle \rho_\uparrow \rho_\downarrow \rangle$ are set to zero since different spins do not correlate if there is no interaction.

These simple results are the basis for the separation of the spin and charge degrees of freedom. This separation is exact in the Random Phase Approximation (RPA), but is only approximate when evaluated with additional terms such as vertex corrections. The RPA is a series of bubble diagrams obtained by evaluating $\langle \rho\rho^\dagger \rangle = P^{(1)}$. In a nonmagnetic system $\langle \rho_c \rho_s^\dagger \rangle = 0$ and the spin and charge interactions do not mix.

For example, now include the interactions $V_{c,s}$ and evaluate the correlations in Eqns. (6.146) and (6.149) using the RPA. In both cases the result is obtained by summing the bubble diagrams

$$P_c(\mathbf{q}, iq_n) = \frac{P^{(1)}(\mathbf{q}, iq_n)}{1 - v_c(\mathbf{q})P^{(1)}(\mathbf{q}, iq_n)} \tag{6.155}$$

$$P_s(\mathbf{q}, iq_n) = \frac{P^{(1)}(\mathbf{q}, iq_n)}{1 + \frac{U}{2}P^{(1)}(\mathbf{q}, iq_n)} \tag{6.156}$$

The denominators of these two expressions are very different. One gives the modes associated with charge fluctuations, while the other the modes from spin fluctuations. In Chapter 5 these types of expressions were evaluated for electrons in energy bands with parabolic dispersion. Let $iq_n \to \omega + i\eta$ where η is infinitesimal. The limits depended on the dimensionless ratio of $x = \omega/(qv_F)$. For large x one had plasmon modes, while for small x one had static screening.

In the present case, evaluate these expressions for energy bands described by a tight-binding model, such as Eq. (6.130). For tight-binding bands define an average velocity $\langle v\rangle = \langle \nabla_k \varepsilon(\mathbf{k})\rangle$. Again define $x = \omega/(q\langle v\rangle)$. The polarization function is

$$P^{(1)}(\mathbf{q},\omega) = \frac{2}{N}\sum_{\mathbf{k}} \frac{n_{\mathbf{k}} - n_{\mathbf{k}+\mathbf{q}}}{\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q}) + \omega + i\eta} \tag{6.157}$$

where the factor of two is for spin summation. Note that the normalization here is slightly different than in Chapter 5: the summation is divided by N rather than by $\mathrm{v} = V_0 N$ where V_0 is the volume of a unit cell in real space.

$\mathbf{x} = \mathbf{0}$: First examine the case that $x = 0$ which is achieved by setting $\omega = 0$. Then the polarization function is static. Next, consider the limit that $\mathbf{q} \to 0$. For tight-binding bands take $\varepsilon(\mathbf{k}+\mathbf{q}) \to \varepsilon(\mathbf{k}) + \mathbf{v}_k \cdot \mathbf{q}$. The latter expression is also used in the occupation function $n_{\mathbf{k}+\mathbf{q}} = n_F(\varepsilon(\mathbf{k}) + \mathbf{v}_k \cdot \mathbf{q}))$ which is expanded for small $\mathbf{q}$

$$\begin{aligned}\lim_{q\to 0} P^{(1)}(\mathbf{q},0) &= \frac{2}{N}\sum_k \frac{n_F[\varepsilon(\mathbf{k})] - [n_F(\varepsilon(\mathbf{k})) + \mathbf{v}_k\cdot\mathbf{q}(dn_F/d\varepsilon(\mathbf{k}))]}{\varepsilon(\mathbf{k}) - [\varepsilon(\mathbf{k}) + \mathbf{v}_k\cdot\mathbf{q}]} \\ &= -2N_f \end{aligned} \tag{6.158}$$

$$N_F = \frac{1}{N}\sum_{\mathbf{k}}\left(-\frac{dn_F(\varepsilon(\mathbf{k}))}{d\varepsilon(\mathbf{k})}\right) \tag{6.159}$$

The factors of $\mathbf{v}_k \cdot \mathbf{q}$ cancel in numerator and denominator, and the final result depends only on the energy derivative of the occupation function. This derivative is a delta function at zero temperature. The density of states for a single spin state is defined as N_F. Using this result in the two susceptibilities (6.155) and (6.156) gives as $\mathbf{q} \to 0$

$$P_c(\mathbf{q},0) = \frac{-2N_F}{1 + N_F[2v(\mathbf{q}) + U]} = \frac{-2N_F^*}{1 + k_s^2/q^2} \tag{6.160}$$

$$P_s = \frac{-2N_F}{1 - UN_F} \tag{6.161}$$

$$N_F^* = \frac{N_F}{1 + UN_F} \tag{6.162}$$

$$k_s^2 = \frac{8\pi e^2 N_F^*}{V_0} \tag{6.163}$$

The Hubbard interaction causes a renormalization of the density of states, for charge fluctuations, to N_F^*. The screening at small wave vector has the Thomas–Fermi form, with a screening wave vector k_s define in Eq. (6.163).

The spin susceptibility P_s has only constants in this limit of $x \to 0$. Its denominator is interesting since it is possible that $1 - UN_F < 0$, which is the *Stoner Criteria* for the onset of a magnetic phase. The phase could either be ferromagnetism or else antiferromagnetism. These possibilities are discussed in the following section. The Stoner criteria assumes that the polarization function $P^{(1)}(\mathbf{q},0)$ has its largest negative value at the center of the Brillouin zone $\mathbf{q} = 0$. However, antiferromagnetism occurs when $1 < UP^{(1)}(\mathbf{q},0)$ for nonzero values of wave vector $\mathbf{q}$. Other interesting choices might have $\mathbf{q}$ at the edge or corner of the Brillouin zone. In this section, it is assumed that the Stoner criteria is not satisfied, and there are no collective magnetic states. Our interest is in the spin and charge fluctuations for the paramagnetic state. Magnetic states are considered in the next section.

Large x: Now consider the other limit where $\omega \gg q\langle v\rangle$. Start with Eq. (6.157) and rewrite it by changing variables $\mathbf{k} \to -\mathbf{q} - \mathbf{k}'$ in the term with $n_{\mathbf{k}+\mathbf{q}}$, which gives

$$P^{(1)}(\mathbf{q},\omega) = \frac{2}{N}\sum_{\mathbf{k}} n_F(\varepsilon(\mathbf{k}))\left[\frac{1}{\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q}) + \omega} + \frac{1}{\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q}) - \omega}\right]$$

Next expand the denominators in a series in inverse powers of the frequency. The expansion parameter is $[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q})]/\omega$. The first nonzero term is

$$\lim_{\omega\to\infty} P^{(1)}(\mathbf{q},\omega) \to -\frac{4}{\omega^2 N}\sum_{\mathbf{k}} n_F(\varepsilon(\mathbf{k}))[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q})] \tag{6.164}$$

The part in square brackets is expanded in powers of $\mathbf{q}$. The first nonzero term is proportional to q^2. In the first-neighbor tight-binding model it is

$$\frac{2}{N}\sum_{\mathbf{k}} n_F(\varepsilon(\mathbf{k}))[\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{q})] = \frac{2t}{N}\sum_{\mathbf{k}} n_F(\varepsilon(\mathbf{k}))\sum_{\boldsymbol{\delta}}(\mathbf{q}\cdot\boldsymbol{\delta})^2 e^{i\mathbf{k}\cdot\boldsymbol{\delta}} = \frac{\delta^2 q^2}{d}\langle -E_K\rangle \tag{6.165}$$

$$\langle -E_K\rangle = \frac{2}{N}\sum_{\mathbf{k}} n_F[\varepsilon(\mathbf{k})]\varepsilon(\mathbf{k}) \tag{6.166}$$

$$P^{(1)}(\mathbf{q},\omega) \to \frac{(\delta q)^2\langle -E_K\rangle}{d\omega^2} \tag{6.167}$$

The summation over $\mathbf{k}$ removes the angular dependence from $\exp(i\mathbf{k}\cdot\boldsymbol{\delta})$ and then the summation over $\boldsymbol{\delta}$ averages $(\mathbf{q}\cdot\boldsymbol{\delta})^2$ to $(q\delta)^2/d$. The symbol d is the dimension. A similar result is obtained for the free electron gas. Note that the average kinetic energy $\langle E_K\rangle$ is a negative number, so $\langle -E_K\rangle$ is positive. This result for $P^{(1)}$ in three dimensions is used in Eqns. (6.155) and (6.156) to give

$$P_c(\mathbf{q},\omega) = \frac{1}{3}\frac{(q\delta)^2\langle -E_K\rangle}{\omega^2 - (\omega_p^2 + c_U^2 q^2)} \tag{6.168}$$

$$\hbar^2\omega_p^2 = \frac{4\pi e^2\delta^2\langle -E_K\rangle}{3V_0} \tag{6.169}$$

$$c_U^2 = \frac{\delta^2 U\langle -E_K\rangle}{6\hbar^2} \tag{6.170}$$

$$P_s(\mathbf{q},\omega) = \frac{1}{3}\frac{(q\delta)^2\langle -E_K\rangle}{\omega^2 + c_U^2 q^2} \tag{6.171}$$

The correlation functions $P_{c,s}$ have denominators which depend upon frequency. Zeros of the denominators represent excitations in the system of particles. The charge fluctuations have their normal behavior of a plasma excitation ω_p at long wavelength. The Hubbard interaction gives the plasmons a quadratic dependence upon wave vector. Note that in the pure Hubbard model, where the long range Coulomb interactions are set to zero, then the excitations in the charge fluctuations become acoustic $\omega = c_U q$. The Hubbard model gets the charge fluctuations wrong, in that it predicts linear waves at long wavelength. The extended Hubbard model, which includes the long range of the Coulomb interaction, gets the correct behavior of a plasma frequency at long wavelength.

The spin fluctuations give damped modes $\omega = \pm iqc_U$. They have very different behavior than the charge fluctuations. These results were derived only using the RPA. However, the general trend of the results is thought to be valid for values of $U \leq W$.

6.3.2. Exchange Graphs

The Hubbard Hamiltonian contains the interaction U multiplied by the combination of $n_{j\uparrow}n_{j\downarrow}$ for every site j. The argument in favor of this form of the interaction is that the exclusion principle forbids two electrons of the same spin from occupying the same orbital on the same site. Only electrons of the opposite spin can occupy the same orbital, and have the interaction U. This feature is built into the Hamiltonian, which assumes the orbital is an s-state which has only spin degeneracy.

There is another way to treat the problem. Instead of building the symmetry into the Hamiltonian, one builds it into the wave function. In this formulation, the Hubbard and extended Hubbard Hamiltonians are

$$H_H = -t\sum_{j\delta\sigma} C^\dagger_{j+\delta,\sigma}C_{j\sigma} + U\sum_{j\sigma\sigma'} n_{j\sigma}n_{j\sigma'} \tag{6.172}$$

$$n_{j\sigma} = C^\dagger_{j\sigma}C_{j\sigma} \tag{6.173}$$

$$H_{xH} = H_H + \frac{1}{2}\sum_{j\neq l,\sigma\sigma'} n_{j\sigma}n_{l\sigma'}V(\mathbf{R}_j - \mathbf{R}_l) \tag{6.174}$$

Now the interactions are between all electrons, regardless of the spin. The feature that no two electrons, with parallel spin, can occupy the same site must be built into the wave function. In our case, it must result from summing the correct sets of diagrams. Which set of diagrams ensure the correct symmetry?

First convert the Hamiltonian to wave vector representation. For the extended Hubbard model

$$H_{xH} = \sum_{\mathbf{k}\sigma} \varepsilon(\mathbf{k})C^\dagger_{\mathbf{k}\sigma}C_{\mathbf{k}\sigma} + \frac{1}{2N}\sum_{\mathbf{q}} v_T(\mathbf{q})\rho_c(\mathbf{q})\rho_c(-\mathbf{q}) \tag{6.175}$$

Note the Coulomb interaction contains only the charge fluctuations, with the full interaction $v_T(\mathbf{q})$. Writing this interaction as $\rho\rho$ allows a particle to interact with itself. This contribution is excluded. Starting from this interaction, the separation of charge and spin is less obvious. Equation (6.175) is called the "full Hubbard model".

The various terms in perturbation theory are examined to find the cancellation which ensures that only electrons of opposite spin occupy and interact on the same lattice site. Generally, the cancellation is done by "exchange graphs". Repeat the discussion in Sec. 5.1. and examine a few terms in the perturbation expansion. In the first order of perturbation theory there are two terms: Hartree and exchange. In the discussion of the jellium model the Hartree term gave no contribution: it utilizes the $\mathbf{q} = 0$ term in the interaction $v_T(0)$, which is absent in jellium. However, here there is a $\mathbf{q} = 0$ term which is U.

$$\Sigma_{H\sigma} = U(n_\sigma + n_{\bar{\sigma}}) \tag{6.176}$$

$$\Sigma_{x\sigma}(\mathbf{p}) = -Un_\sigma - \frac{1}{N}\sum_{\mathbf{q}} v(\mathbf{q})n_{\mathbf{p}+\mathbf{q}} \tag{6.177}$$

$$\Sigma_{H\sigma} + \Sigma_{x\sigma} = Un_{\bar{\sigma}} - \frac{1}{N}\sum_{q} v(\mathbf{q})n_{\mathbf{p}+\mathbf{q},\sigma} \tag{6.178}$$

where σ is the spin of the electron, while $\bar{\sigma}$ signifies the opposite spin. The symbols n_σ denote the average concentration. A paramagnetic system is assumed, which has $n_{\bar{\sigma}} = n_\sigma$. The total first-order self-energy, including Hartree and exchange, is $Un_{\bar{\sigma}}$ for the Hubbard

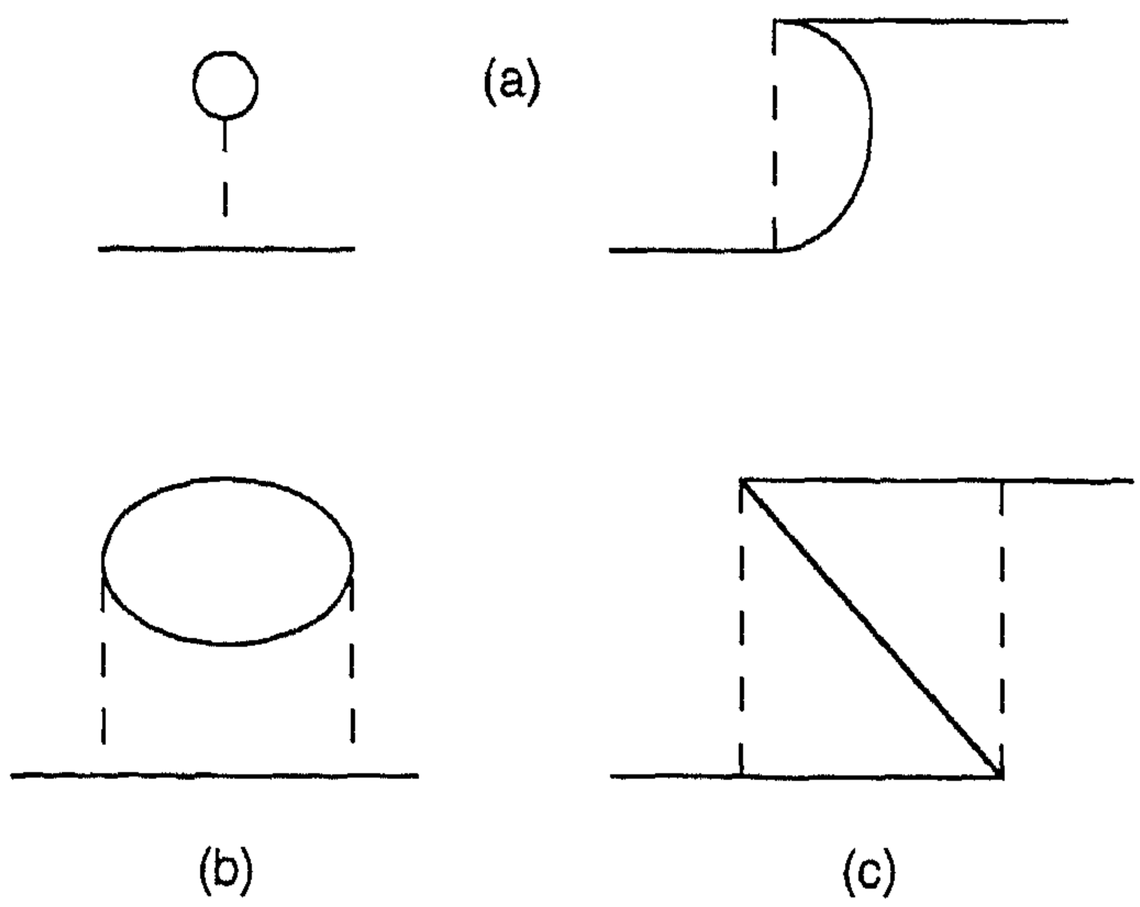


FIGURE 6.9

model, which is precisely what one gets from the usual Hubbard Hamiltonian with the interaction $U\sum_j n_{j\sigma}n_{j\bar{\sigma}}$. The extended Hubbard model also has an exchange term of the usual form for charge fluctuations. The same cancellation occurs in the second order of perturbation theory. The RPA part of the self-energy is the diagram in Fig. 6.9(b), while Fig. 6.9(a) shows Hartree and exchange. Figure 6.9(b) is a closed loop connected to the Green's function by two interactions. Let P_σ denote the closed loop polarization diagram for electrons of spin σ. In the usual Hubbard model, the closed loop must have opposite spin of the electron, so it gives a self-energy term with an effective interaction of $U^2P_{\bar{\sigma}}$. In the usual extended Hubbard model, the interaction $v(\mathbf{q})$ can interact with a closed loop of either spin. The self-energy in second order, from the usual extended Hubbard model, is

$$\Sigma_\sigma^{(2)}(p) = \frac{1}{N}\sum_{\mathbf{q}} \mathcal{G}(p+q)P_{\bar{\sigma}}(q)[2v(\mathbf{q})^2 + 2Uv(\mathbf{q}) + U^2] \tag{6.179}$$

The term with $2Uv(\mathbf{q})$ comes from Fig. 6.9(b) where the two interaction lines have U on one side and $v(\mathbf{q})$ on the other: the factor of two since there are two ways of doing this. For the full Hubbard model in (6.175), with all spin interactions, the same diagram gives

$$\Sigma_\sigma^{(2)}(p)' = \frac{1}{N}\sum_{\mathbf{q}} \mathcal{G}(p+q)P_{\bar{\sigma}}(q)[2v(\mathbf{q})^2 + 4Uv(\mathbf{q}) + 2U^2] \tag{6.180}$$

which is different. Including the exchange diagram in Fig. 6.9(c) makes these two self-energies more alike.

$$\begin{aligned}\Sigma_{x\sigma}^{(2)}(p) &= -\frac{1}{N^2}\sum_{qq'} v_T(\mathbf{q})v_T(\mathbf{q}')\mathcal{G}(p+q)\mathcal{G}(p+q')\mathcal{G}(p+q+q') \\ &= -\frac{1}{N}\sum_{q} \mathcal{G}(p+q)P_{\bar{\sigma}}(q)[U^2 + 2Uv(\mathbf{q})] \\ &\quad -\frac{1}{N^2}\sum_{qq'} v(\mathbf{q})v(\mathbf{q}')\mathcal{G}(p+q)\mathcal{G}(p+q')\mathcal{G}(p+q+q')\end{aligned}$$

Adding the two self-energy terms gives

$$\Sigma_\sigma^{(2)}(p) = \frac{1}{N}\sum_{\mathbf{q}} \mathcal{G}(p+q)P_{\bar{\sigma}}(q)[2v(\mathbf{q})^2 + 2Uv(\mathbf{q}) + U^2]$$
$$- \frac{1}{N^2}\sum_{qq'} v(\mathbf{q})v(\mathbf{q}')\mathcal{G}(p+q)\mathcal{G}(p+q')\mathcal{G}(p+q+q') \tag{6.182}$$

The first term is identical to (6.179). The exchange graph Fig. 6.9(c) cancels out the second-order parallel spin terms in the RPA term (Fig. 6.9(b)) for the Hubbard model. Note that the exchange graphs also produce exchange contributions from the charge fluctuations.

The RPA expansion for the extended Hubbard model treats the interaction U differently from $v(\mathbf{q})$. The exchange graphs must be included in order to cancel out the parallel spin parts of the U interation. When doing RPA for the charge fluctuation interaction $v(\mathbf{q})$ the exchange contributions are usually omitted. Starting from (6.175) demonstrates this difference. This different treatment may not cause problems. The exchange graphs from charge fluctuations give large contributions to the energy of the electrons, but contribute little to the dynamics: collective modes, etc.

6.4. HUBBARD MODEL: MAGNETIC PHASES

The Hubbard model has various phases with magnetic ordering. The usual metallic phase, with no magnetic ordering, is called *paramagnetic* and is abbreviated "P" on phase diagrams. It has equal number of electrons with spin up and down in the various bands, and the spins have no ordering except local correlation. They have short-range but not long-range order among the spin alignments. Other magnetic phases which are considered here are: (i) *ferromagnetic*, where there is a net magnetic moment in some direction, which is caused by having a majority of conduction electrons with spins pointing in the same direction. It is denoted as "F" in phase diagrams (ii) *antiferromagnetic* ordering where there are an equal number of conduction electrons with spins up and with spin down. The symbol "A" is used for this phase. The distinction between PM and AF is that in AF the spins alignments have both long-range order and short-range order. These three phases will be discussed in great detail below.

Other magnetic orderings are possible. A *charge-density wave* is denoted CDW. Its signature is that the average value of the charge density operator

$$\langle \rho(\mathbf{q})\rangle = \sum_{\mathbf{k}\sigma} \langle C^\dagger_{\mathbf{k}+\mathbf{q},\sigma} C_{\mathbf{k}\sigma}\rangle \tag{6.183}$$

has a value of $O(N)$ for some nonzero value of $\mathbf{q}$. CDWs usually occur when the vector $\mathbf{q}$ connects two regions of the Fermi surface, which is called *nesting*. Similarly, one form of *spin-density waves* (SDW) is given by

$$\langle \rho_s(\mathbf{q})\rangle = \sum_{\mathbf{k}} [\langle C^\dagger_{\mathbf{k}+\mathbf{q},\uparrow} C_{\mathbf{k}\uparrow}\rangle - \langle C^\dagger_{\mathbf{k}+\mathbf{q},\downarrow} C_{\mathbf{k}\downarrow}\rangle] \tag{6.184}$$

The existence of CDW and SDW was first suggested by Overhauser (1964, 1978), and they have been found in many materials. They are not discussed here. Some work has suggested that the Hubbard model has regions of its phase space where there exist CDW and SDW.

The ground state energy of the Hubbard model will be solved for the various magnetic phases. Depending upon the variables, each of the three phases may have the lowest energy,

and be the ground state. What are the important variables? There are two energy scales: the Hubbard U and the bandwidth $W = 2zt$ of the kinetic energy, where z is the coordination number. Usually the results are plotted according to the ratio of these two variables, such as the combination U/W. The inverse ratio is used sometimes. Another important variable is the concentration n of conduction electrons on the lattice. In the usual Hubbard model, the band is full when $n = 2$. This case is uninteresting since none of the electrons can move and the system is an insulator. The lattice has an average of one electron per site when $n = 1$. This case is important since the AF phase occurs when $n \approx 1$. The temperature is an important variable, since the magnetic phases disappear at high temperature. In two dimensions, they disappear at any nonzero temperature, since the Mermin–Wagner theorem (1966) states there is never long-range order in 2D.

The Hubbard model has *particle hole symmetry*, which means that the phase diagram is symmetric. It has reflection symmetry around the point $n = 1$. Any phase that exists for $0 < n < 1$ also exists for $n \to 2 - n$. This simple theorem is easy to prove. The reference state for electron calculations is the quasiparticle vacuum $|0\rangle_e$. We define another reference state $|FB\rangle$ which is the state if the electronic band is completely full of electrons: the state with $n = 2$. The symbol "FB" denotes "full band". It is easy to calculate the energy of this state in the real space representation. Since all states are full, the kinetic energy term $C^\dagger_{j+\delta} C_j|FB\rangle = 0$. The state on $j + \delta$ is already occupied and the creation operator gives zero. In real space, all site occupancies $n_{j\sigma}$ are one, so the Hubbard interaction is UN. The result is $H|FB\rangle = NU|FB\rangle$.

A *hole* is defined as an electron missing from the state $|FB\rangle$. Let $|0\rangle_h$ denote the "hole vacuum" and the hole operators are $d, d^\dagger$. Then

$$C_{j\sigma}|FB\rangle = d^\dagger_{j\bar{\sigma}}|0\rangle_h \tag{6.185}$$

$$C^\dagger_{j\sigma}|FB\rangle = d_{j\bar{\sigma}}|0\rangle_h \tag{6.186}$$

Destruction of an electron is equivalent to the creation of a hole. The spin changes sign between the electron and hole operator. The destruction of an electron of spin up leaves the band with net spin down, which is assigned to the hole. Using the commutator for Fermions, the number operator for electrons is changed to a number operator for holes

$$C^\dagger_{j\sigma} C_{j\sigma}|FB\rangle = d_{j\bar{\sigma}} d^\dagger_{j\bar{\sigma}}|0\rangle_h \tag{6.187}$$

$$= [1 - d^\dagger_{j\bar{\sigma}} d_{j\bar{\sigma}}]|0\rangle_h \tag{6.188}$$

$$H_e = H_h + UN(1 - n) \tag{6.189}$$

$$H_h = t \sum_{j\delta\sigma} d^\dagger_{j+\delta,\sigma} d_{j\sigma} + U \sum_j n^{(h)}_{j\sigma} n^{(h)}_{j\bar{\sigma}} \tag{6.190}$$

$$n^{(h)}_{j\sigma} = d^\dagger_{j\sigma} d_{j\sigma} \tag{6.191}$$

The hole Hamiltonian H_h has nearly the same form as the original electron Hamiltonian H_e. The difference is that the hopping term t has changed sign. This inverts the bands, which is no problem. In the square lattice we can select as the center of the Brillouin zone the wave vector $\mathbf{Q} = (\pi, \pi)/a$. Since $\cos(\theta_x + Q_x) = -\cos(\theta_x)$, this choice of band center inverts the bands back to their original form. An alternate, but equivalent procedure, is to define the hole operators with a minus sign ($C_{j\sigma}|FB\rangle = -d^\dagger_{j\bar{\sigma}}|0\rangle_h$) on one of the two sublattices, which then retains the negative sign in front of the hopping term. The hole problem is identical to the electron problem. Any electron state of n has the same hole state of $2 - n$. This proves

particle–hole symmetry. The constant terms in the transformation, $NU(1-n)$, are phase independent so do not influence the selection of the magnetic ordering. This discussion completes the digression into the explanation of particle–hole symmetry.

Another important variable for magnetic ordering is to choose the lattice for the calculation. The results depend upon the choice of the lattice, and also depend upon whether the calculation is done in two or three dimensions. The early work on the Hubbard model concentrated on three dimensions, and most calculations were done for the simple-cubic (sc) or the body-centered-cubic (bcc) structures. The discovery of high temperature superconductivity in the copper oxides, called "cuprates", heightened interested in layered compounds. This focussed attention on the two dimensional lattice. Most recent work has been done on the square lattice (sq) in two dimensions. This lattice has the advantage that the Brillouin Zone is a plane and is represented nicely on a sheet of paper. The sq lattice has coordination number $z = 4$ and $W = 8t$. The hopping energy, from the interaction with nearest neighbors, is

$$\varepsilon(\mathbf{k}) = -2t[\cos(\theta_x) + \cos(\theta_y)] \tag{6.192}$$

$$\theta_\mu = k_\mu a \tag{6.193}$$

where a is the lattice constant.

6.4.1. Ferromagnetism

In thinking about the *ferromagnetic state*, it is useful to keep in mind two theorems. The earliest was by Kanamori (1963), who proved that in the limit of large U, a system which is doped slightly away from half filling is always ferromagnetic. The second theorem, by Vilk *et al.* (1994), is that the the Hubbard model on the sq lattice is never ferromagnetic. These two theorems directly disagree, and show that the Hubbard model is still not completely understood.

The ferromagnetic state has a majority of spins pointing along one axis. This axis can be in any direction, since the theory is isotropic. Take it to be spin up. Denote the concentrations of up spin as n_+ and down spins as n_-. Use the symbol n_σ where $\sigma = \pm$. Also, $n_{\bar{\sigma}}$ is the density of spins opposite to σ: the overbar denotes a negative value. The net concentration is $n = n_+ + n_-$. Ferromagnetism occurs whenever $n_+ \neq n_-$.

Denote by m the fractional magnetization.

$$m = \frac{n_+ - n_-}{n_+ + n_-} \tag{6.194}$$

Values of the magnetization range $-1 \leq m \leq 1$. The spin densities may be expressed as

$$n_+ = \frac{n}{2}(1+m) \tag{6.195}$$

$$n_- = \frac{n}{2}(1-m) \tag{6.196}$$

The Hubbard model was solved earlier using RPA. The first-order self energy was from the Hartree interaction, and gives that $E_\sigma(\mathbf{k}) = \varepsilon(\mathbf{k}) + Un_{\bar{\sigma}}$. In this approximation the density of up spins is

$$n_\sigma = \frac{1}{N}\sum_{\mathbf{k}} n_{\mathbf{k}\sigma} \tag{6.197}$$

$$= \int \frac{d^2\theta}{(2\pi)^2} \frac{1}{e^{\beta(\varepsilon(\mathbf{k})+Un_{\bar{\sigma}}-\mu)}+1} \tag{6.198}$$

The θ integrals range from $-\pi \leq \theta_\mu \leq \pi$ for the sq lattice.

Let $\rho(\varepsilon)$ be the density of states. Equation (6.194) can be used to derive a self-consistent equation for the magnetization at zero temperature. Then the occupation number becomes a step function $n_F(\varepsilon - \mu^*) \to \Theta(\mu^* - \varepsilon)$

$$\rho(\varepsilon) = \int \frac{d^2\theta}{(2\pi)^2} \delta(\varepsilon - \varepsilon(\mathbf{k})) \tag{6.199}$$

$$m = \frac{1}{n}\int d\varepsilon \rho(\varepsilon)[\Theta(\mu^* - \varepsilon + mnU/2) - \Theta(\mu^* - \varepsilon - mnU/2)]$$

$$\mu^* = \mu - \frac{n}{2}U \tag{6.200}$$

$$n = \int d\varepsilon \rho(\varepsilon)[\Theta(\mu^* - \varepsilon + mnU/2) + \Theta(\mu^* - \varepsilon - mnU/2)] \tag{6.201}$$

The renormalized chemical potential μ^* is determined by the concentration n as given in Eq. (6.201).

The phase boundary is the point where the magnetization begins: it is the line where there is a solution for the above equation with $m = 0^+$. Expand the right-hand side of the equation for m in a power series in the small parameter m. This side is proportional to m in this limit. Cancelling the factor of m from both sides of the equation gives the equation which determines the onset of ferromagnetism at zero temperature

$$1 = U\rho(\mu^*) \tag{6.202}$$

Equation (6.202) is the first one to predict a phase boundary. There are several caveats. The first is that the equation is very approximate. Only Hartree theory is used in its derivation, which is neither sophisticated nor accurate. Secondly, this same result earlier was derived earlier: it is just the Stoner criteria found in the prior section using RPA. The Stoner criteria was given in terms of the polarization operator $P^{(1)}(\mathbf{q}, \omega)$ in the case that $\omega = 0$ while $\mathbf{q} \to 0$

$$\lim_{\mathbf{q}\to 0}\left\{1 + \frac{U}{2}P^{(1)}(\mathbf{q}, 0)\right\} = 0 \tag{6.203}$$

At zero temperature, in the indicated limit the polarization diagram is exactly equal to $-2\rho(\mu^*)$. Then the Stoner criteria is identical to Eq. (6.202). It is interesting that the simple theory is identical to RPA.

The next step is to evaluate the density of states, which is given in terms of an elliptic function. First, define a dimensionless energy $x = \varepsilon/(4t)$ and dimensionless state density $\tilde{\rho}(x)$

$$\rho(\varepsilon) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi} d\theta_x \int_{-\pi}^{\pi} d\theta_y \delta[\varepsilon + 2t(\cos(\theta_x) + \cos(\theta_y))] \tag{6.204}$$

$$= \frac{1}{4t}\tilde{\rho}(x) \tag{6.205}$$

$$\tilde{\rho}(x) = \frac{2}{4\pi^2}\int_{-\pi}^{\pi} d\theta_x \int_{-\pi}^{\pi} d\theta_y \delta[2x + \cos(\theta_x) + \cos(\theta_y)] \tag{6.206}$$

$$\tilde{\rho}(x) = \frac{2}{\pi^2} K(\sqrt{1 - x^2}) \tag{6.207}$$

The elliptic function $K(z)$ diverges when $z = 1$, which in our case is when $x = 0$. This point turns out to be half-filling: when $n = 1$. The resulting density of states is shown in Fig. 6.10. Note that the dimensionless energy spans the range of $-1 \leq x \leq 1$. The band is half-full at $x = 0$ where the density of states diverges to infinity at the van Hove singularity. The divergence is logarthmic and the density of states is integrable through this divergence. In fact, the total density of states integrates to unity:

$$1 = \int_{-1}^{1} dx \tilde{\rho}(x) \tag{6.208}$$

which makes sense since $\rho(\varepsilon)$ is the density of states for a single spin configuration. The other energy variables are also normalized to $4t$: $u = U/4t$, $\nu = \mu/4t$.

It may be possible to have a ferromagnetic state with partial magnetization $0 < |m| < 1$. The onset of the FM state is defined as $m \neq 0$. However, all the examples we have solved have the feature that at the phase boundary m jumps to unity. Therefore, that will be assumed here,

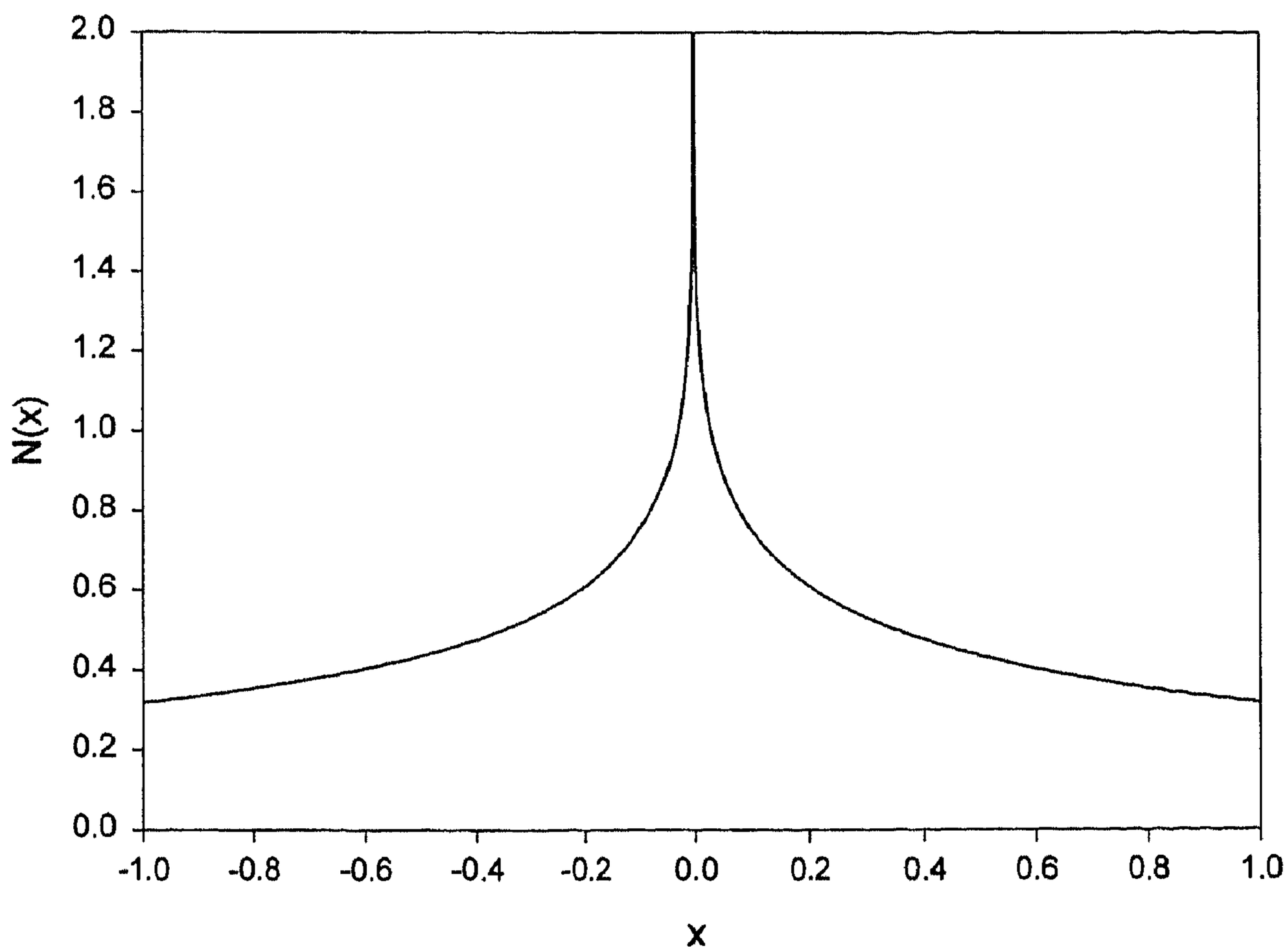


FIGURE 6.10 Density of states for an electron on a square lattice.

in order to simplify the algebra. In the PM state $m = 0$ and in the FM state $|m| = 1$. In this case if $n_\sigma = n$ and $n_{\bar{\sigma}} = 0$, the chemical potential for the majority spins is just $\nu^* = \nu$. The feature that $n_{\bar{\sigma}} = 0$ requires that $\nu - un < -1$, which can be shown from the definition

$$n_{\bar{\sigma}} = \int_{-1}^{1} dx\tilde{\rho}(x)\Theta(\nu - un - x) \tag{6.209}$$

The integral is zero if the argument of the theta function is never positive.

The Stoner criteria for the phase boundary between FM and PM is found by solving together the two equations

$$n = \int_{-1}^{1} dx\tilde{\rho}(x)\Theta(\nu - x) \tag{6.210}$$

$$u_c = \left(\frac{U_c}{4t}\right) = \frac{1}{\tilde{\rho}(\nu)} \tag{6.211}$$

The first equation (6.210) is solved to give $n(\nu)$, where ν is the dimensionless chemical potential which ranges $-1 \leq \nu \leq 1$. The second equation (6.211) is solved to give the phase boundary $(U_c/4t)$ also as a function of ν. This result is combined with the first calculation to give the phase boundary $(U_c/4t)$ as a function of n. The predicted phase boundary reaches $U_c = 0$ at $n \approx 1$, which occurs because of the divergence in the density of states for the sq lattice at half filling. Most other lattices do not have this divergence, and at $n = 1$ ferromagnetism only occurs above a critical value of the Hubbard parameter U_c. The sq lattice is unusual in having this phase boundary at $U_c = 0$ at $n = 1$. All of this discussion is actually pointless, since at $n = 1$ the system prefers to be in the antiferromagnetic state for small values of U, rather than in the ferromagnetic state.

6.4.2. Antiferromagnetism

Figure 6.11(a) shows a lattice of circles and squares. If all of the points were alike, this structure is the square lattice. The area of the unit cell is one square. In the AF state, the circles form one sublattice, while the squares form another. Antiferromagnetism occurs when the circle sublattice has all electrons with spin up, while the square sublattice has all electrons with spin down. Or vice versa. In real space, the unit cell in the antiferromagnetic state is the square formed by four nearest neighbor circles. This cell has exactly twice the area of the original unit cell. The occurrence of antiferromagnetism, in this case, doubles the size of the unit cell. If A_0 is the area of the cell in real space, and A_{BZ} is the area of the Brillouin zone, then the theorem of areas states that $(2\pi)^2 = A_0 A_{BZ}$. When one doubles the area in real space, the area of the Brillouin zone reduces in half.

Figure 6.11(b) shows the Brillouin zone for the sq lattice. The units assume a bond length of $a = 1$ so the minimal reciprocal lattice vector is $\mathbf{G} = 2\pi(\pm 1, 0)$ or $2\pi(0, \pm 1)$. The Brillouin zone for the sq lattice is the large square whose edges are solid lines. In the AF state, the Brillouin zone is half as large. Then its boundaries are the dashed lines in the shape of a diamond. The dashed line is also the Fermi surface in the paramagnetic state for half-filling $(n = 1)$. This line has an energy $\varepsilon(\mathbf{k}) = 0$. Also shown in the figure is the vector $\mathbf{Q} = (\pi, \pi)$ which spans the Fermi surface at half-filling. Another similar wave vector $\mathbf{Q}' = (\pi, -\pi)$ is not shown, but it spans the Fermi surface in the direction perpendicular to $\mathbf{Q}$. The vector $\mathbf{Q} \pm \mathbf{Q}'$ equals a reciprocal lattice vector, so that $\mathbf{Q}$ and $\mathbf{Q}'$ are regarded as equivalent. They play a

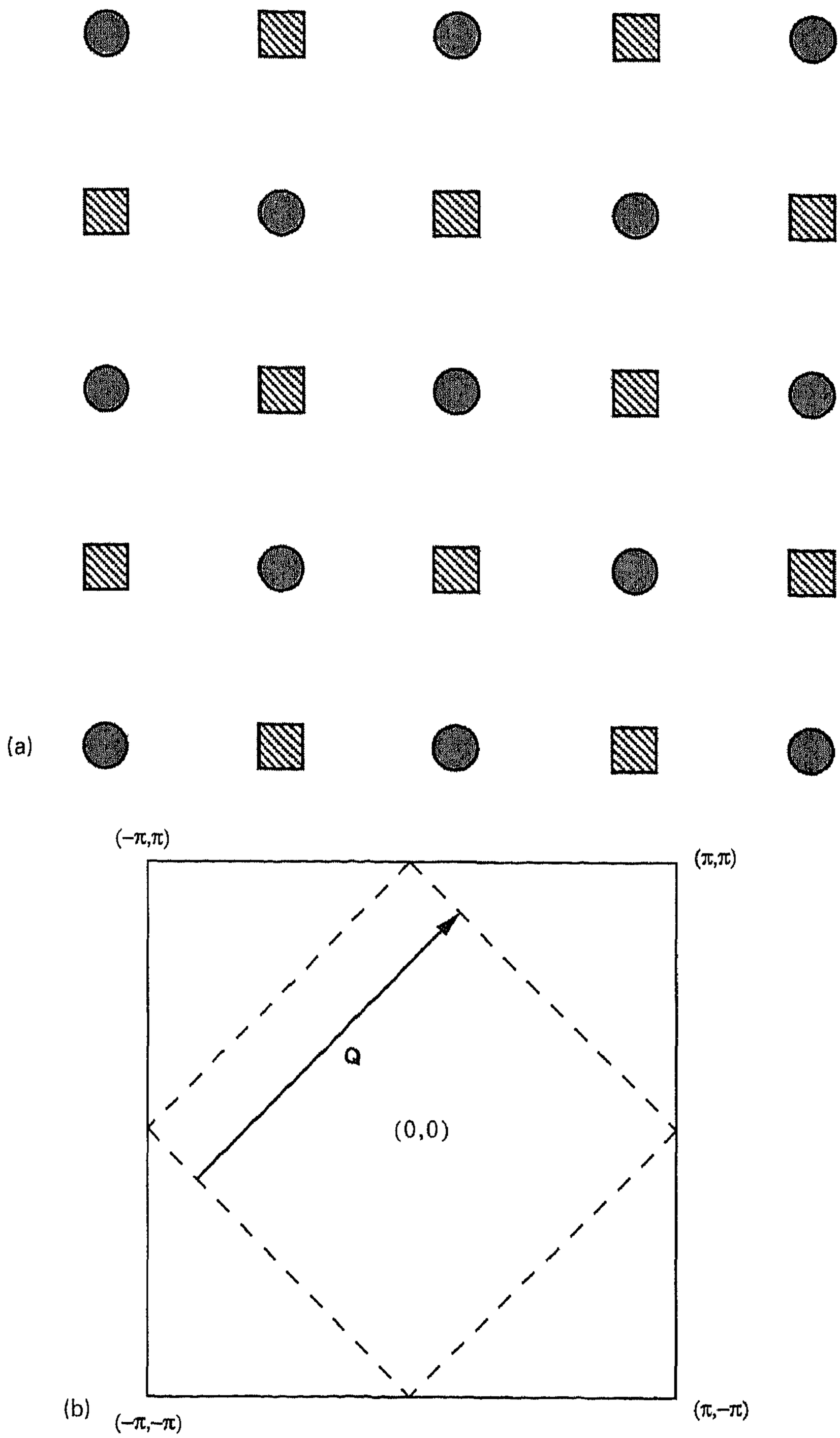


FIGURE 6.11 The square lattice. (a) In AF phase, circles have electrons of spin up, squares have electrons of spin down. (b) Brillouin zone for sq lattice.

special role in the theory of antiferromagnetism. From now on the symbol **Q** is used to denote either of the four vectors $(\pm\pi, \pm\pi)$.

Let **k** denote any wave vector in the Brillouin zone for the AF state. These are wave vectors in the diamond shaped area in Fig. 6.11(b). The combination of $\mathbf{k} + \mathbf{Q}$ takes one to a state outside of this reduced zone, but still within the original Brillouin zone of the para-

magnetic state. In a reduced zone representation of the AF state, the wave vector points $\mathbf{k}$ and $\mathbf{k}+\mathbf{Q}$ are the same point. In the PM state they are different points. So the onset of antiferromagnetism is governed by order parameters of the form

$$A_m(\mathbf{Q}) = \frac{1}{N}\sum_{\mathbf{k}}[\langle C^\dagger_{\mathbf{k}+\mathbf{Q},\sigma} C_{\mathbf{k},\sigma}\rangle - \langle C^\dagger_{\mathbf{k}+\mathbf{Q},\bar{\sigma}} C_{\mathbf{k},\bar{\sigma}'}\rangle] \tag{6.212}$$

$$A_{\sigma\bar{\sigma}}(\mathbf{Q}) = \frac{1}{N}\sum_{\mathbf{k}}\langle C^\dagger_{\mathbf{k}+\mathbf{Q},\sigma} C_{\mathbf{k},\bar{\sigma}}\rangle \tag{6.213}$$

where N is the number of lattice sites. These order parameters are similar to those for charge and spin density waves. The AF state is a special kind of SDW. Either of the above two order parameters can be used to describe the correlation in the AF state. The feature that a single wave vector $\mathbf{Q}$ connects large segments of the Fermi surface is an example of nesting.

The sq lattice is a typical lattice for the description of antiferromagnetism. It naturally divides into two sublattices, which doubles the unit cell in real space and halves the Brillouin zone. There are other lattices which have a different behavior. Figure 6.12(a) shows the honeycomb (hc) lattice which has two different lattice sites in each unit cell. Each site has three nearest neighbors. Antiferromagnetism occurs when one site has electrons with spin σ while the other site has electrons with spin $\bar{\sigma}$. In this case the unit cell does not change size, and the Brillouin zone in the AF state is identical to that of the PM state. Then the correlation

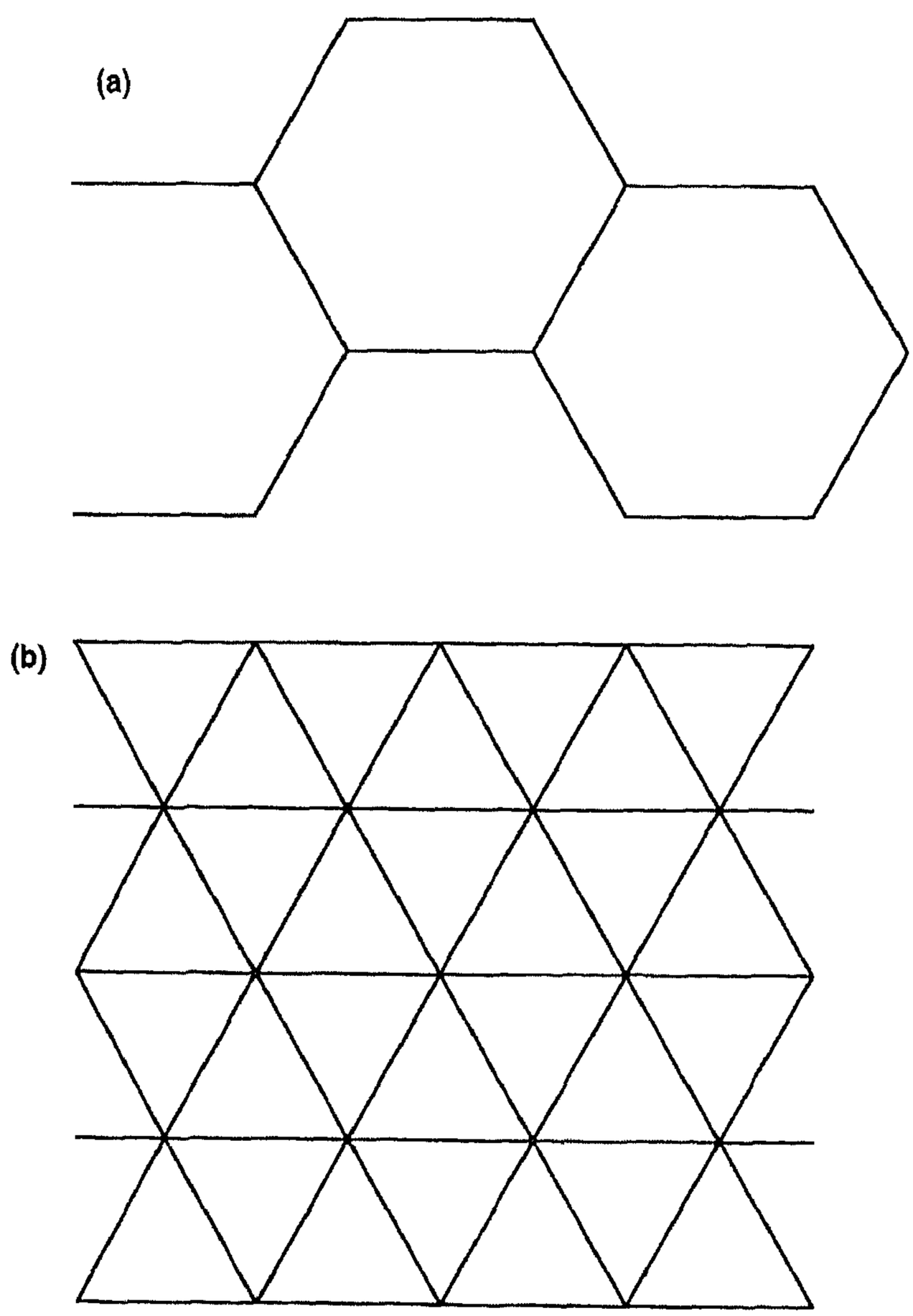


FIGURE 6.12 (a) The hc lattice. (b) The pt lattice.

functions for the onset of the AF state have $\mathbf{Q} = 0$. The hc lattice occurs in nature as the structure of a single plane of graphite, where the vertices are carbon atoms. In graphite the chemical σ-bonds are the lines in the plane and are fully occupied. The conduction band is composed of p_z orbitals which are perpendicular to the plane. They bond isotropically to the three first neighbors, and the band structure is given accurately by the first neighbor tight-binding model with an overlap parameter $t = 2.4$ eV. The band is half full ($n = 1$). There is no experimental evidence for an antiferromagnetic state, probably because U is too small. Both the sq and hc lattices are *bipartite*, which means they have an obvious way of supporting an antiferromagnetic structure with alternate up and down spins.

Figure 6.12(b) shows the plane triangular lattice (pt). The vertices are lattice sites, while the lines show bond directions. Each lattice point has six first neighbors. The pt lattice is not bipartite. The basic structure in the unit cell is the equilateral triangle. This figure cannot support antiferromagnetism: if one vertex has spin up, and another has spin down, then the third has an equal tendency to be up or down, and is *frustrated*. It is believed that the pt lattice does not have an antiferromagnetic structure at zero magnetic field at half-filling. All three of these lattices (sq,hc,pt) support a ferromagnetic structure.

Now consider the antiferromagnetic transition for the square lattice. Start by examining the function

$$A_{\mathbf{q}\sigma} = \frac{1}{N}\sum_{\mathbf{k}}\langle C^{\dagger}_{\mathbf{k}+\mathbf{q},\sigma} C_{\mathbf{k}\sigma}\rangle \tag{6.214}$$

$$= \frac{1}{N}\sum_{j}\langle n_{j\sigma}\rangle e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{6.215}$$

The first expression is the charge density wave correlation function expressed in wave vector space. The second equation is the same expression expressed in real space. For the paramagnetic state, each site j has equal probability of having an up or down spin. In this case $\langle n_{j\sigma}\rangle = n/2$. Then the correlation function equals zero unless $\mathbf{q} = 0$.

A different result is obtained in the antiferromagnetic state. Then one sublattice has all spins σ while the other sublattice has all spins $\bar{\sigma}$. If one sublattice has spin up, and the other has spin down, then

$$A_m(\mathbf{q}) = A_{\mathbf{q}\uparrow} - A_{\mathbf{q}\downarrow} \tag{6.216}$$

$$= \frac{n}{N}\sum_{\text{every other } R_j} e^{i\mathbf{q}\cdot\mathbf{R}_j}[1 - e^{i\mathbf{q}\cdot\boldsymbol{\tau}}] \tag{6.217}$$

$$\boldsymbol{\tau} = (\pm 1, 0), \text{ or } (0, \pm 1) \tag{6.218}$$

The largest value of the above correlation function is when $\exp(i\mathbf{q}\cdot\boldsymbol{\tau}) = -1$ which is when $\mathbf{q} = \mathbf{Q}$. This is another argument for why the wave vector $\mathbf{Q}$ is special.

Bogoliubov theory is used for the calculation of the AF state. Start with the standard Hubbard Hamiltonian

$$H = \sum_{\mathbf{k}\sigma}\varepsilon(\mathbf{k})n_{\mathbf{k}\sigma} + \frac{U}{N}\sum_{\mathbf{kpq}} C^{\dagger}_{\mathbf{p}+\mathbf{q}\uparrow} C^{\dagger}_{\mathbf{k}-\mathbf{q}\downarrow} C_{\mathbf{k}\downarrow} C_{\mathbf{p}\uparrow} \tag{6.219}$$

In the Hartree approximation for the paramagnetic state, evaluate the Hubbard energy at $\mathbf{q} = 0$ to get two energy terms

$$\frac{U}{N}\sum_{\mathbf{pk}}[\langle C^\dagger_{\mathbf{p}\uparrow}C^\dagger_{\mathbf{p}\uparrow}\rangle C^\dagger_{\mathbf{k}\downarrow}C_{\mathbf{k}\downarrow} + \langle C^\dagger_{\mathbf{k}\downarrow}C_{\mathbf{k}\downarrow}\rangle C^\dagger_{\mathbf{p}\uparrow}C_{\mathbf{p}\uparrow}] \tag{6.220}$$

$$= U\sum_{\mathbf{k}}[n_\uparrow n_{\mathbf{k}\downarrow} + n_\downarrow n_{\mathbf{k}\uparrow}] \tag{6.221}$$

$$H = \sum_{\mathbf{k}\sigma}[\varepsilon(\mathbf{k}) + Un_{\bar\sigma}]n_{\mathbf{k}\sigma} - UNn_\sigma n_{\bar\sigma} \tag{6.222}$$

The last term in the effective Hamiltonian is a constant $-UNn_\sigma n_{\bar\sigma}$ which is included to get the correct total energy. The Bogoliubov approximation overcounts the interaction, so this term is subtracted to get the correct value. For the antiferromagnetic state, include the correlations with $\mathbf{q} = \mathbf{Q}$. In this case the pairing of the last term in (6.219) is

$$\frac{U}{n}\sum_{\mathbf{pk}}[\langle C^\dagger_{\mathbf{p}+\mathbf{Q}\uparrow}C_{\mathbf{p}\uparrow}\rangle C^\dagger_{\mathbf{k}-\mathbf{Q}\downarrow}C_{k\downarrow} + \langle C^\dagger_{\mathbf{k}-\mathbf{Q}\downarrow}C_{\mathbf{k}\downarrow}\rangle C^\dagger_{\mathbf{p}+\mathbf{Q}\uparrow}C_{\mathbf{p}\uparrow}]$$

The brackets are $A_{\mathbf{Q}\sigma}$. The effective Hamiltonian in the Bogoliubov approximation is

$$H = \sum_{\mathbf{k}\sigma}[\varepsilon(\mathbf{k}) + Un_{\bar\sigma}]n_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma}[\varepsilon(\mathbf{k}+\mathbf{Q}) + Un_{\bar\sigma}]n_{\mathbf{k}+\mathbf{Q}\sigma}$$
$$+ U\sum_{\mathbf{k}\sigma}A_{\mathbf{Q}\bar\sigma}[C^\dagger_{\mathbf{k}+\mathbf{Q}\sigma}C_{\mathbf{k}\sigma} + C^\dagger_{\mathbf{k}\sigma}C_{\mathbf{k}+\mathbf{Q}\sigma}] - UN(n_\sigma n_{\bar\sigma} + A_{\mathbf{Q}\sigma}A_{\mathbf{Q}\bar\sigma}) \tag{6.223}$$

Constant terms are again subtracted from the effective Hamiltonian to get the correct ground state energy. The summation over $\mathbf{k}$ is now only for the Brillouin zone of the AF state. The states $\mathbf{k}+\mathbf{Q}$ are added in order to sum over all of the states in the PM zone. For the sq lattice $\varepsilon(\mathbf{k}+\mathbf{Q}) = -\varepsilon(\mathbf{k})$. This Hamiltonian is easily diagonalized by using the transformation

$$\alpha_{\mathbf{k}\sigma} = \cos(\theta_k)C_{\mathbf{k}\sigma} + \sin(\theta_k)C_{\mathbf{k}+\mathbf{Q}\sigma} \tag{6.224}$$

$$\beta_{\mathbf{k}\sigma} = -\sin(\theta_k)C_{\mathbf{k}\sigma} + \cos(\theta_k)C_{\mathbf{k}+\mathbf{Q}\sigma} \tag{6.225}$$

$$C_{\mathbf{k}\sigma} = \cos(\theta_k)\alpha_{\mathbf{k}\sigma} - \sin(\theta_k)\beta_{\mathbf{k}\sigma} \tag{6.226}$$

$$C_{\mathbf{k}+\mathbf{Q}\sigma} = \sin(\theta_k)\alpha_{\mathbf{k}\sigma} + \cos(\theta_k)\beta_{\mathbf{k}\sigma} \tag{6.227}$$

The two operators $C_{\mathbf{k}\sigma}$ and $C_{\mathbf{k}+\mathbf{Q}\sigma}$ are replaced by $\alpha_{\mathbf{k}\sigma}$ and $\beta_{\mathbf{k}\sigma}$. The angle θ_k is determined below.

This transformation is inserted into the Hamiltonian (6.223). Some algebra gives the expression below. We omit writing the subscripts $(\mathbf{k}\sigma)$ on each term.

$$H = \sum_{\mathbf{k}\sigma}\{[\alpha^\dagger\alpha + \beta^\dagger\beta]Un_{\bar\sigma} + [\alpha^\dagger\alpha - \beta^\dagger\beta][\varepsilon\cos(2\theta_k) + UA_{\mathbf{Q}\bar\sigma}\sin(2\theta_k)]$$
$$+ (\alpha^\dagger\beta + \beta^\dagger\alpha)[UA_{\mathbf{Q}_{\bar\sigma}}\cos(2\theta_k) - \varepsilon\sin(2\theta_k)]\} \tag{6.228}$$

Since α and β are to be pure eigenstates, the last term in the above Hamiltonian must vanish. It vanishes by selecting the angle θ_k to be

$$\tan(2\theta_k) = \frac{UA_{\mathbf{Q}\bar{\sigma}}}{\varepsilon} \tag{6.229}$$

$$E(\mathbf{k}) = \sqrt{\varepsilon(\mathbf{k})^2 + (UA_{\mathbf{Q}\bar{\sigma}})^2} \tag{6.230}$$

$$\begin{aligned} H = &\sum_{\mathbf{k}\sigma}\{\alpha^\dagger_{\mathbf{k}\sigma}\alpha_{\mathbf{k}\sigma}[Un_{\bar{\sigma}} + E(\mathbf{k})] + \beta^\dagger_{\mathbf{k}\sigma}\beta_{\mathbf{k}\sigma}[Un_{\bar{\sigma}} - E(\mathbf{k})]\} \\ &- UN(n_\sigma n_{\bar{\sigma}} + A_{\mathbf{Q}\sigma}A_{\mathbf{Q}\bar{\sigma}}) \end{aligned} \tag{6.231}$$

The energy spectrum has developed a gap which is twice $UA_{\mathbf{Q}\bar{\sigma}}$. The dimensionless gap is $\Delta = uA_{\mathbf{Q}\sigma}$. The energy $E(\mathbf{k})$ is similar to that found in the BCS theory of superconductivity in Chapter 10. There are some important differences. The main one is that the chemical potential μ does *not* go into the expression for $E(\mathbf{k})$: note it is not written as $\sqrt{(\varepsilon - \mu)^2 + (UA_{\mathbf{Q}\bar{\sigma}})^2}$. This difference is significant since the energy gap is only important if there are regions of occupied electron states where $\varepsilon(\mathbf{k}) = 0$. In the sq lattice, there are such energy states along the dashed-diamond shaped curve in Fig. 6.12(b). However, if the density of electrons n is changed away from $n = 1$ then these zero-energy states may not be occupied. The fact they could have a gap becomes less important. Then the tendency for the AF state diminishes. The lack of the chemical potential in $E(\mathbf{k})$ is significant, and makes the theory much different from the BCS theory of superconductivity.

As in BCS theory, in Chapter 10, the final step in the calculation is to obtain a self-consistent equation for the order parameter $A_{\mathbf{Q}\sigma}$. Again restricting the summations over wave vector to the Brillouin zone in the AF state, and using the transformation to the states (α, β)

$$A_{\mathbf{Q}\sigma} = \frac{1}{N}\sum_{\mathbf{k}}[\langle C^\dagger_{\mathbf{k}+\mathbf{Q}\sigma}C_{\mathbf{k}\sigma}\rangle + \langle C^\dagger_{\mathbf{k}\sigma}C_{\mathbf{k}+\mathbf{Q}\sigma}\rangle] \tag{6.232}$$

$$= \frac{1}{N}\sum_{\mathbf{k}}\sin(2\theta_k)[\alpha^\dagger_{\mathbf{k}\sigma}\alpha_{\mathbf{k}\sigma} - \beta^\dagger_{\mathbf{k}\sigma}\beta_{\mathbf{k}\sigma}] \tag{6.233}$$

$$A_{\mathbf{Q}\sigma} = A_{\mathbf{Q}\bar{\sigma}}\frac{U}{N}\sum_{\mathbf{k}}\frac{1}{E(\mathbf{k})}[n_F(E(\mathbf{k}) - \mu^*) - n_F(-E(\mathbf{k}) - \mu^*)] \tag{6.234}$$

The β state has the negative energy with respect to the Fermi surface, and it is the state which is most occupied at low temperatures. The expression in brackets in the last equation, which is the difference in occupation numbers for the alpha and beta states, is negative. The solution to this equation requires that $A_{\mathbf{Q}\sigma} = -A_{\mathbf{Q}\bar{\sigma}}$. This choice is in agreement with Eq. (6.212). The maximum order parameter occurs when $A_{\mathbf{Q}\sigma} = -A_{\mathbf{Q}\bar{\sigma}}$. The self-consistent equation for the energy gap in the AF state is

$$1 = \frac{U}{N}\sum_{\mathbf{k}}\frac{1}{E(\mathbf{k})}[n_F(-E(\mathbf{k}) - \mu^*) - n_F(E(\mathbf{k}) - \mu^*)] \tag{6.235}$$

$$= \frac{U}{N}\sum_{\mathbf{k}}\frac{1}{E(\mathbf{k})}\frac{\sinh(\beta E(\mathbf{k}))}{\cosh(\beta\mu^*) + \cosh(\beta E(\mathbf{k}))} \tag{6.236}$$

At zero temperature the expression is written as

$$1 - \frac{U}{N}\sum_{\mathbf{k}}\frac{1}{E(\mathbf{k})}[\Theta(E(\mathbf{k}) + \mu^*) - \Theta(\mu^* - E(\mathbf{k}))] \tag{6.237}$$

Again let the density of states in the sq lattice be $\rho(\varepsilon)$. Recast this equation as

$$1 = U\int_{-4t}^{0} d\varepsilon \frac{\rho(\varepsilon)}{E(\varepsilon)} \frac{\sinh(\beta E)}{\cosh(\beta\mu^*) + \cosh(\beta E)} \tag{6.238}$$

$$E(\varepsilon) = \sqrt{\varepsilon^2 + (UA_{\mathbf{Q}\sigma})^2} \tag{6.239}$$

Note that the upper limit of integration is now zero, since the summation over $\mathbf{k}$ only extended over the reduced Brillouin zone. The effective chemical potential $\mu^* = \mu - Un_{\bar{\sigma}}$ is determined by requiring that the number of particles remain n

$$n = \frac{1}{N}\sum_{\mathbf{k}\sigma} [n_{\mathbf{k}\sigma} + n_{\mathbf{k}+\mathbf{Q}\sigma}] \tag{6.240}$$

$$= \frac{1}{N}\sum_{\mathbf{k}\sigma} [\alpha^\dagger_{\mathbf{k}\sigma}\alpha_{\mathbf{k}\sigma} + \beta^\dagger_{\mathbf{k}\sigma}\beta_{\mathbf{k}\sigma}] \tag{6.241}$$

$$= \int_{-4t}^{0} d\varepsilon\rho(\varepsilon) \frac{\cosh(\beta E)}{\cosh(\beta\mu^*) + \cosh(\beta E)} \tag{6.242}$$

If $\mu^* = 0$ then the right-hand side integrates to one, as it should. This identity is valid regardless of the value of the energy gap. For values of n away from $n = 1$ one has to solve this equation in order to obtain $n(\mu^*)$, which therefore gives $\mu^*(n)$.

The gap equation is (6.238). It is identical to the Stoner criteria. Recall that the Stoner criteria for the AF state is given in terms of the polarization function $P_\sigma(\mathbf{Q}, 0)$ for a single spin component

$$1 = -UP_\sigma(\mathbf{Q}, 0) \tag{6.243}$$

$$P_\sigma(\mathbf{Q}, 0) = \frac{1}{N}\sum_{\mathbf{k}} \frac{n_k - n_{k+Q}}{\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{Q})} \tag{6.244}$$

$$= \frac{2}{N}\sum_{\mathbf{k}} \frac{n_k}{\varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}+\mathbf{Q})} \tag{6.245}$$

The formula simplifies when using the feature that $\varepsilon(\mathbf{k}+\mathbf{Q}) = -\varepsilon(\mathbf{k})$.

$$P_\sigma(\mathbf{Q}, 0) = \frac{1}{N}\sum_{\mathbf{k}} \frac{n_k}{\varepsilon(\mathbf{k})} \tag{6.246}$$

$$= \int_{-4t}^{4t} d\varepsilon \frac{\rho(\varepsilon)}{\varepsilon} n_F(\varepsilon - \mu^*) \tag{6.247}$$

$$= -\int_{-4t}^{0} d\varepsilon \frac{\rho(\varepsilon)}{\varepsilon} \frac{\sinh(\beta\varepsilon)}{\cosh(\beta\mu^*) + \cosh(\beta\varepsilon)} \tag{6.248}$$

Here the summations over k extend over all of the Brillouin zone of the PM state. The Stoner criteria is valid only at the transition temperature where the gap is zero. In this case the Stoner theory is identical to (6.238) derived using the Bogoliubov approximation.

6.4.3. An Example

It is useful to give an example of these results. A simple model is best, in which all results are analytical. To this end assume that the density of states $\rho(\varepsilon)$ is a simple constant

over the range of energy $-W \leq \varepsilon \leq W$ and is zero elsewhere. The normalization which integrates to one electron is

$$\rho(\varepsilon) = \begin{cases} \dfrac{1}{2W} & -W \leq \varepsilon \leq W \\ 0 & |\varepsilon| \geq W \end{cases} \tag{6.249}$$

It is convenient to use dimensionless units such as $x = \varepsilon/W$, $\nu = \mu/W$, and $u = U/W$. Assume zero temperature so that Fermi functions become step functions.

In the *paramagnetic state*, then $n_\sigma = n_{\bar{\sigma}} = n/2$. Define $\mu^* = \mu - Un/2$ and $\nu^* = \nu - un/2$. The density of particles is

$$n = 2\int_{-W}^{\mu^*} d\varepsilon\rho(\varepsilon) = 1 + \nu^* \tag{6.250}$$

The factor of two is for spin degeneracy. Define the reduced ground state energy per electron as $\mathscr{E}_j = E_G/(NW)$ where the symbol j denotes P,F, or A for our three possible states. For the PM state

$$E_G = \sum_{\mathbf{k}\sigma}\left[\varepsilon(\mathbf{k}) + U\frac{n}{2}\right]n_{\mathbf{k}\sigma} - UN\left(\frac{n}{2}\right)^2 \tag{6.251}$$

$$\mathscr{E}_P = \frac{2}{W}\int_{-W}^{\mu^*} d\varepsilon\left[\varepsilon + U\frac{n}{2}\right]\rho(\varepsilon) - \frac{U}{W}\left(\frac{n}{2}\right)^2 \tag{6.252}$$

$$= \tfrac{1}{2}[(v^*)^2 - 1] + (n\mu^* + 1)u\frac{n}{2} - \frac{un^2}{4} \tag{6.253}$$

$$\mathscr{E}_P = \tfrac{1}{2}n(n-2) + \frac{un^2}{4} \tag{6.254}$$

In Eq. (6.251) the last term is subtracted since the Hubbard term is taken twice in the kinetic energy part, where it is overcounted. In the last expression (6.254), $\nu^* = (n-1)$ is used to obtain an expression for the reduced ground state energy as a function of occupation number n and the Hubbard parameter ($u = U/W$).

The assumption in the *ferromagnetic state* is that the up (σ) and down ($\bar{\sigma}$) spins have unequal occupation. Assume that up is dominent $n_\sigma > n_{\bar{\sigma}}$. Again define the magnetization as $m = (n_\sigma - n_{\bar{\sigma}})/n$. An interesting result is obtained by separately calculating these two occupation numbers by integrating over the density of states

$$n_\sigma = \int_{-W}^{\mu - Un_\sigma} d\varepsilon\rho(\varepsilon) = \tfrac{1}{2}[u + 1 - un_{\bar{\sigma}}] \tag{6.255}$$

$$n_{\bar{\sigma}} = \int_{-W}^{\mu - Un_\sigma} d\varepsilon\rho(\varepsilon) = \tfrac{1}{2}[u + 1 - un_\sigma] \tag{6.256}$$

Subtract these two equations, and the difference in the occupations is the magnetization, which gives the interesting result

$$nm = nm\frac{u}{2} \tag{6.257}$$

This equation has two solutions for $n \neq 0$: either $m = 0$ or else $u = 2$. The latter is the Stoner criteria: $1 = U\rho(0) = U/(2W)$. What these equations actually mean is that there is no solution for $m \neq 0$ until $u \geq 2$. At this point then m jumps to one. The onset of the FM state is given by the inequality $U > 2W$. The calculation of the ground state energy is rather easy.

Since $n_{\bar{\sigma}} = 0$ there is no Hubbard energy: all the spins are alike to avoid any Hubbard energy. The ground state occupation and energy for the FM state are

$$n_\sigma = n = \tfrac{1}{2}(1 + \nu) \tag{6.258}$$

$$E_G = N \int_{-W}^{\mu} d\varepsilon \varepsilon \rho(\varepsilon) \tag{6.259}$$

$$\mathscr{E}_F = \tfrac{1}{4}(\nu^2 - 1) = n(n-1) \tag{6.260}$$

where the expression for $\nu(n)$ is used to obtain the last expression. The corrrectness of this result can be confirmed by comparing the ground state energies for the FM and PM states. The FM state has the lowest energy whenever $\mathscr{E}_F < \mathscr{E}_P$ which gives

$$n(n-1) < \tfrac{1}{2}n(n-2) + \frac{un^2}{4} \tag{6.261}$$

which has the solution that $2 < u$ for all values of occupation n. The phase diagram for a system with just two ordering possibilities, PM or FM, is quite simple. The boundary is a horizontal line at $U/W = 2$. Above the line is the FM phase and below it is the PM phase. The line is flat because a constant density of states is assumed. The curve is actually a plot of $U(n) = 1/\rho[\mu(n)]$. If the density of states is not constant, then the phase boundary has curvature.

The third phase is *antiferromagnetic* (*AF*). First evaluate the Stoner criteria for the phase boundary between the AF and the PM states

$$1 = -U \int_{-W}^{\mu} \frac{d\varepsilon}{\varepsilon} \rho(\varepsilon) = -\frac{u}{2} \ln|\nu| \tag{6.262}$$

$$U_c(n) = -\frac{2W}{\ln|n-1|} \tag{6.263}$$

This expression has three singular points. At $n = 0, 2$ then $\ln|n-1| = \ln(1) = 0$ and the value for U_c diverges to positive infinity. The third point is at $n = 1$. There one has $\ln(0)$ which diverges to negative infinity, so that $U_c(1) = 0$. This behavior gives the phase boundary shown in Fig. 6.13(a). There are listed three phases where they exist alone. In the top-middle region there could exist either AF or PM. The ground state energy of the AF state must be calculated in order to compare the ground state energies of these two phases. The calculation of the properties of the AF state proceeds in several steps, since there are several quantities to determine. The first is the relationship between the occupation number n and the reduced chemical potential ν^*. The relevant expression at zero temperature is

$$n = \int_{-1}^{0} dx[\Theta(\nu^* - \sqrt{x^2 + \Delta^2}) + \Theta(\nu^* + \sqrt{x^2 + \Delta^2})] \tag{6.264}$$

$$\Delta = uA_{Q\sigma} \tag{6.265}$$

Recall that the energy is $E = \sqrt{\varepsilon^2 + (UA_{Q\sigma})^2}$ which when divided by W becomes $\sqrt{x^2 + \Delta^2}$. The upper limit of integration is zero, rather than one, since the integral is over the reduced Brillouin zone which only goes halfway in energy. The two theta functions are the Fermi–Dirac occupation functions for the α_k and β_k operators. First consider the case that $n < 1$ so that $\nu^* < 0$. Then the first theta function is always zero, while the second gives $\nu^* + \sqrt{x^2 + \Delta^2} > 0$ which can be solved for negative x to give

$$-\sqrt{(\nu^*)^2 - \Delta^2} > x \tag{6.266}$$

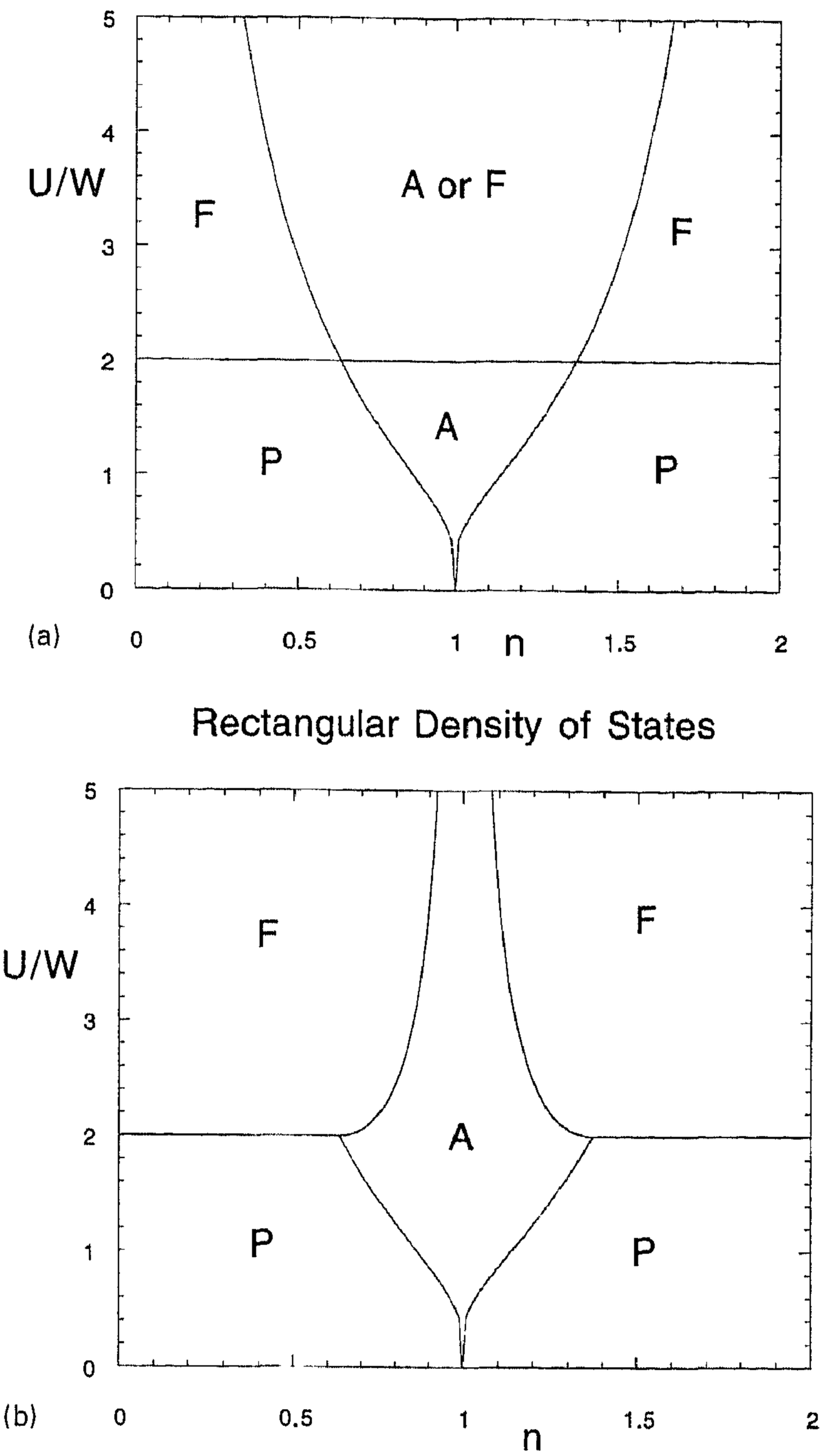


FIGURE 6.13 (a) Preliminary phase diagram for the Hubbard model. Region with AF and FM together is to be determined. (b) Final phase diagram for the Hubbard model.

so that the integral for the occupation number gives

$$n = 1 - \sqrt{(\nu^*)^2 - \Delta^2} \tag{6.267}$$

$$\nu^* = -\sqrt{\Delta^2 + (1-n)^2} \tag{6.268}$$

This equation provides one relation between ν^*, Δ and n. A similar derivation for $n > 1$ gives

$$n = 1 + \sqrt{(\nu^*)^2 - \Delta^2} \tag{6.269}$$

$$\nu^* = \sqrt{\Delta^2 + (1-n)^2} \tag{6.270}$$

The second step in the derivation of the AF state is to solve the integral (6.237) for the gap parameter Δ. In dimensionless units it is the equation

$$1 = \frac{u}{2}\int_{-1}^{0} \frac{dx}{\sqrt{x^2+\Delta^2}}[\Theta(\nu^* + \sqrt{x^2+\Delta^2}) - \Theta(\nu^* - \sqrt{x^2+\Delta^2})]$$

$$= \frac{u}{2}\ln\left(x + \sqrt{x^2+\Delta^2}\right)_{-1}^{n-1} \tag{6.271}$$

$$= \frac{u}{2}\ln\left(\frac{\sqrt{\Delta^2+(n-1)^2} - (1-n)}{\sqrt{\Delta^2+1} - 1}\right) \tag{6.272}$$

The upper limit of the x integral for $\nu^* < 0$ is $(n-1)$, as shown in the derivation of the occupation number. Solve the above equation by defining $S = \exp(2/u)$, and then exponentiate the equation. Also define $y = \sqrt{1+\Delta^2}$

$$\sqrt{\Delta^2+(n-1)^2} = S(y-1) + (1-n) \tag{6.273}$$

After squaring this equation, there is a quadratic equation for y, which is solved to obtain

$$y = \frac{-S(1-n-S) + |S(1-n)-1|}{S^2-1} \tag{6.274}$$

If the factor of $S(1-n)-1$ is positive, then $y = 1$ and $\Delta = 0$. In this case there is no AF state. This result is expected, since the criteria that $S(1-n) > 1$ is equivalent to $u < -2/\ln(1-n)$. From the Stoner criteria, the system is not in the AF state for this small value of u. For $S(1-n) < 1$ there is a solution with $y > 1$ which gives the order parameter

$$y = \sqrt{\Delta^2+1} = \coth(2/u) - (1-n)\mathrm{csch}(2/u) \tag{6.275}$$

$$\Delta^2 = \left\{\coth(1/u) - \frac{(1-n)}{\sinh(1/u)}\right\}^2 - 1 \tag{6.276}$$

$$\sqrt{\Delta^2+(n-1)^2} = \mathrm{csch}(2/u) - (1-n)\coth(1/u) \tag{6.277}$$

This nice result depends only upon n and u. It can be evaluated, and then the result used to find ν^* in Eq. (6.268).

The third and final step in solving the properties of the AF state is to evaluate the integral for the ground state energy, again at zero temperature. In dimensionless units it is the integral in (6.231)

$$\mathscr{E}_A = \int_{-1}^{0} dx\left\{\left[\frac{un}{2} + \sqrt{x^2+\Delta^2}\right]\Theta\left(\nu^* - \sqrt{x^2+\Delta^2}\right)\right.$$
$$\left. + \left[\frac{un}{2} - \sqrt{x^2+\Delta^2}\right]\Theta\left(\nu^* + \sqrt{x^2+\Delta^2}\right)\right\} - u(n/2)^2 + \Delta^2/u \tag{6.278}$$

The constant term Δ^2/u has changed sign since $A_{Q\bar{\sigma}} = -A_{Q\sigma}$. The integral can be evaluated easily by changing integration variables to ξ where $x = -\Delta\sinh(\xi)$ and $\sqrt{x^2+\Delta^2} = \Delta\cosh(\xi)$.

$$\int_{-1}^{n-1} dx\sqrt{x^2+\Delta^2} = \Delta^2\int_{\xi_n}^{\xi_1} d\xi\cosh^2\xi \tag{6.279}$$

$$= \frac{\Delta^2}{2}\left[\xi + \tfrac{1}{2}\sinh(2\xi)\right]_{\xi_n}^{\xi_1} \tag{6.280}$$

$$= \Delta^2/u + \tfrac{1}{2}[\sqrt{1+\Delta^2} - (1-n)\sqrt{(1-n)^2+\Delta^2}]$$

$$1 - n = \Delta\sinh(\xi_n) \tag{6.281}$$

$$1 = \Delta\sinh(\xi_1) \tag{6.282}$$

Collecting these results gives

$$\varepsilon_A = \frac{u}{4}n^2 + \tfrac{1}{2}[(1-n)\sqrt{)1-n)^2+\Delta^2} - \sqrt{(1-n)^2+\Delta^2} - \sqrt{(1-n)^2+\Delta^2} - \sqrt{1-\Delta^2}] \tag{6.283}$$

$$\varepsilon_A = \frac{u}{4}n^2 + \tfrac{1}{2}\left[\frac{2(1-n)}{\sinh(1/u)} - [2-n(2-n)]\coth(1/u)\right] \tag{6.284}$$

The first expression is the direct result of the prior algebra. The second equation is found by using Eq. (6.276) for Δ^2 and doing more algebra. It has the advantage that the ground state energy is a direct function of the variables (n, u). This expression is the final result needed for the calculation of the phase diagram in the Bogoliubov approximation. For each value of n and u, one can calculate the order parameter Δ and also the ground state energy per electron $\mathscr{E}_A$. For $u > 2$ the AF energy is compared to that of the FM state $[\mathscr{E}_F = n(n-1)]$ which is independent of u. The resulting phase diagram is shown in Fig. 6.13(b). For $u > 2$ the FM state occupies an increasingly large part of the phase space, for different values of n. The region of AF state gets shrunk to an increasingly small region near to $n \approx 1$. The reader is cautioned that these results are quite approximate, and are based solely on the Bogoliubov approximation to the initial Hamiltonian. Probably the AF phase does not extend down to zero value of U at $n = 1$, but instead there is a critical value of U_c for the onset of the AF state. This feature is discussed in the following section.

Figure 6.13(b) shows why the Kanamori theorem is perhaps reasonable. As U becomes large the FM phase approaches the line $n = 1$ on both sides. His theorem is that in the limit $U \to \infty$, then a bipartite lattice with N sites will be ferromagnetic if the number of electrons differs slightly from N.

6.4.4. Local Field Corrections

The homogeneous electron gas was discussed in Chapter 5. There the RPA was found to be a poor approximation for most quantities. It was necessary to add some additional correlation into the response functions, by postulating the existence of Hubbard local field

corrections: $G_H(q)$ for charge fluctuations, and a similar function for spin fluctuations. A similar approach to the Hubbard model would suggest the RPA result is changed to

$$\chi_c(q) = \frac{P^{(1)}(q)}{1 - g_c \frac{U}{2} P^{(1)}(q)} \tag{6.285}$$

$$\chi_s(q) = \frac{P^{(1)}(q)}{1 + g_s \frac{U}{2} P^{(1)}(q)} \tag{6.286}$$

where $P^{(1)}(q)$ is the usual polarization function of the electronic system, as calculated using a single bubble diagram without vertex or other corrections. A better evaluation of the polarization diagram would include vertex corrections. The present idea is to call these vertex corrections the constants $g_{c,s}$, which just multiply the correlation function. These vertex constants will be a function of U/W. This simple idea follows directly from the Hubbard approach. The question is how to calculate these vertex constants.

A simple and clever way was suggested by Vilk and Tremblay (1997). They used sum rules on the correlation functions to provide formulas for the vertex constants. These sum rules are derived from the exact definitions

$$\chi_c(\mathbf{q}, iq_n) = -\frac{1}{N}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau \rho_c(\mathbf{q}, \tau)\rho_c(-\mathbf{q}, 0)\rangle \tag{6.287}$$

$$\chi_s(\mathbf{q}, iq_n) = -\frac{1}{N}\int_0^\beta d\tau e^{iq_n\tau}\langle T_\tau \rho_s(\mathbf{q}, \tau)\rho_s(-\mathbf{q}, 0)\rangle \tag{6.288}$$

The sum rules are derived by taking a four-vector summation over these correlation functions: both over wave vector $\mathbf{q}$ and imaginary frequency q_n. First do the summation over frequency. This step simplifies the expression, since the summation over frequency gives a delta function $\delta(\tau)$

$$\frac{1}{\beta}\sum_{q_n} e^{iq_n\tau} = \delta(\tau) \tag{6.289}$$

$$\frac{1}{\beta}\sum_{q_n} \chi_c(\mathbf{q}, iq_n) = -\frac{1}{N}\langle \rho_c(\mathbf{q}, 0)\rho_c(-\mathbf{q}, 0)\rangle \tag{6.290}$$

$$\frac{1}{\beta}\sum_{q_n} \chi_s(\mathbf{q}, iq_n) = -\frac{1}{N}\langle \rho_s(\mathbf{q}, 0)\rho_s(-\mathbf{q}, 0)\rangle \tag{6.291}$$

The next step is to write the density operators in real space such as

$$\rho_c(\mathbf{q}) = \sum_j [n_{j\uparrow} + n_{j\downarrow}]e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{6.292}$$

$$\rho_s(\mathbf{q}) = \sum_j [n_{j\uparrow} - n_{j\downarrow}]e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{6.293}$$

$$\frac{1}{N}\sum_{\mathbf{q}}\langle \rho_c(\mathbf{q})\rho_c(-\mathbf{q})\rangle = \sum_j [n_{j\uparrow} + n_{j\downarrow}]^2 \tag{6.294}$$

$$\frac{1}{N}\sum_{\mathbf{q}}\langle \rho_s(\mathbf{q})\rho_s(-\mathbf{q})\rangle = \sum_j [n_{j\uparrow} - n_{j\downarrow}]^2 \tag{6.295}$$

These expressions are evaluated using $n_{j\sigma}^2 = n_{j\sigma}$ to finally get

$$n_\sigma = \frac{1}{N}\sum_j \langle n_{j\sigma}\rangle \tag{6.296}$$

$$\langle n_\uparrow n_\downarrow\rangle = \frac{1}{N}\sum_j \langle n_{j\uparrow} n_{j\downarrow}\rangle \tag{6.297}$$

$$\frac{1}{\beta N}\sum_{\mathbf{q},q_n} \chi_c(\mathbf{q}, iq_n) = -[n + 2\langle n_\uparrow n_\downarrow\rangle] \tag{6.298}$$

$$\frac{1}{\beta N}\sum_{\mathbf{q},q_n} \chi_s(\mathbf{q}, iq_n) = -[n - 2\langle n_\uparrow n_\downarrow\rangle] \tag{6.299}$$

These two results are close to the final formula for the sum rule. The one for the spin correlation is in its final form. A change must be made in the formula for the charge correlation. In the expression for the charge fluctuations, there is a contribution to the sum rule which is omitted in RPA. This contribution is the term with $\mathbf{q} = 0$. Then the density operator is $\rho(\mathbf{q}, \tau) = N_e$, which is independent of τ. This term gives a contribution which is $N_e^2 \delta_{\mathbf{q}=0}\delta_{n=0}$. It only contributes at zero frequency and at zero wave vector. It is usually omitted when considering fluctuations. However, this term is included in the sum rule which sums over all $\mathbf{q}$ and all n. Since it is omitted from RPA, also omit it from the sum rule. Rewrite the sum rule for the charge correlation to be

$$\frac{1}{\beta N}\sum_{\mathbf{q}q_n} \chi_c(\mathbf{q}, iq_n) = -[n + 2\langle n_\uparrow n_\downarrow\rangle - n^2] \tag{6.300}$$

The two sum rules are Eqs. (6.299) and (6.300). The derivation is unfinished without an algorithm for determining the important correlation $\langle n_\uparrow n_\downarrow\rangle$. Vilk *et al.* (1994) argued that it was just $g_s = \langle n_\uparrow n_\downarrow\rangle/(\langle n_\uparrow\rangle\langle n_\downarrow\rangle)$. The sum rule for the spin correlation is now just a self-consistent nonlinear equation for $\langle n_\uparrow n_\downarrow\rangle$. It is determined first. Afterwards, the charge fluctuation sum rule can be evaluated to give g_c. In this way the two vertex constants $g_{c,s}$ are found self-consistently. Results of Vilk and Tremblay for the two-dimensional sq lattice at half-filling ($n = 1$) are shown in Fig. 6.14. They plot the effective Hubbard constants $U_c = g_c U$, $U_s = g_s U$ as a function of U, where the hopping energy $t = 1$. The charge interaction U_c is significantly increased by the vertex interactions, while the spin interaction

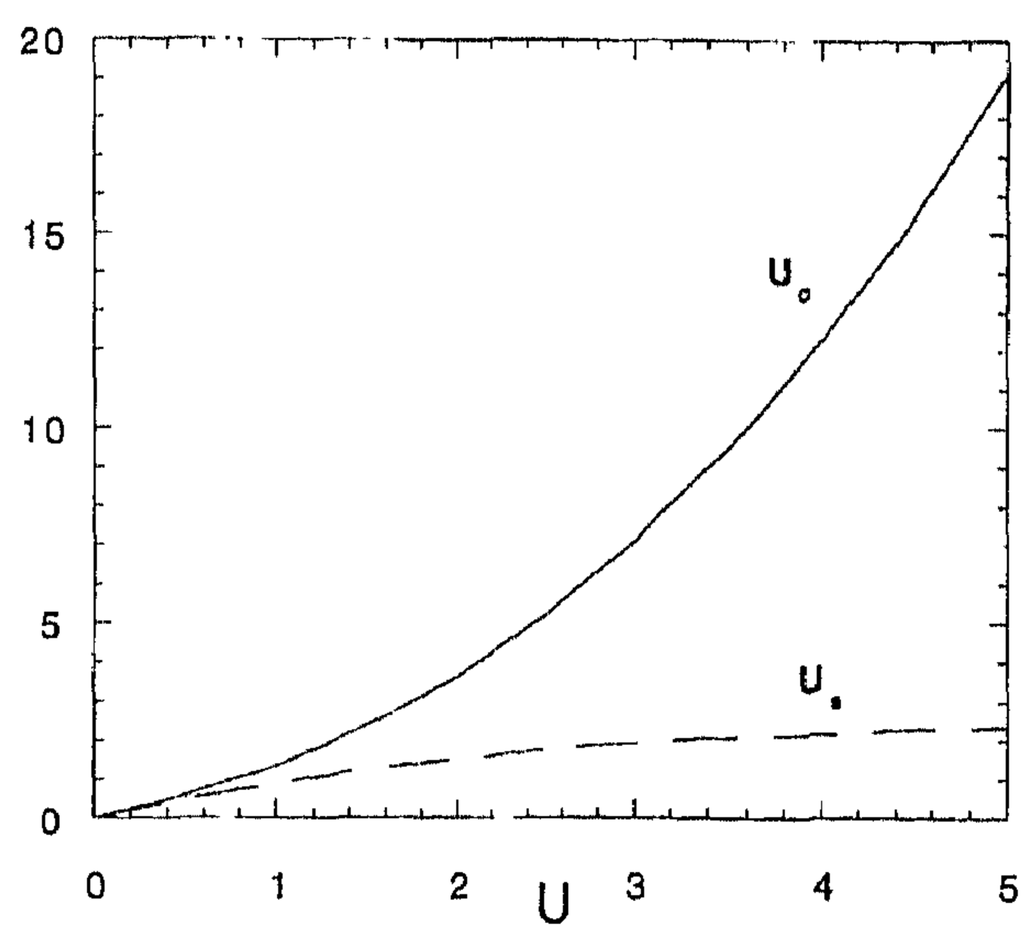


FIGURE 6.14 U_c and U_s for the Hubbard model on the sq lattice as a function of U at half-filling ($n = 1$). From Vilk and Tremblay (1997) (used with permission).

U_s is significantly decreased. Similar changes occur at other values of filling n. Obviously, this change in the effective interaction changes the locations of the phase boundaries.

Their theorem about the lack of an FM phase is based upon these types of calculations. As $U \to \infty$ the effective U_s saturates at a small value, which is less than needed for the onset of the FM phase.

They also calculate the self-energy of the electron including the scattering by charge and spin fluctuations

$$\Sigma_\sigma(p) = Un_{\bar{\sigma}} - \frac{U}{4\beta N}\sum_q [U_c\chi_c(q) + U_s\chi_s(q)]\mathscr{G}^{(0)}(p+q) \tag{6.301}$$

where a four-vector notation is used, as usual. The summation is over wave vector $\mathbf{q}$ and frequency q_n. Note that only one vertex is dressed, so that the interaction with the charge fluctuations is UU_c, rather than U_c^2. The latter would be overcounting the vertex corrections. They find this simple analytic self-energy function agrees quite well with numerical results obtained using quantum Monte Carlo simulations.

Problems

1. Consider the bound states of the Kondo Hamiltonian for a localized spin one-half particle interacting with a single particle in a band of continuum states:

$$H = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}\sigma} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{k}\sigma} - \sum_{\mathbf{k}\mathbf{k}'} [J_z S^{(z)}(C^\dagger_{\mathbf{k}\uparrow}C_{\mathbf{k}'\uparrow} - C^\dagger_{\mathbf{k}\downarrow}C_{\mathbf{k}'\downarrow}) + J_\perp(C^\dagger_{\mathbf{k}\uparrow}C_{\mathbf{k}'\downarrow}S^{(-)} + C^\dagger_{\mathbf{k}\downarrow}C_{\mathbf{k}k\uparrow}S^{(+)}] \tag{6.302}$$

where $S^{(z)}$, $S^{(-)}$ and $S^{(+)}$ refer to the localized spins. Show that the following are exact bound states of H:

$$\phi_1 = \sum_{\mathbf{k}} a_{\mathbf{k}} C^\dagger_{\mathbf{k}\uparrow}|\uparrow\rangle \tag{6.303}$$

$$\phi_2 = \sum_{\mathbf{k}} b_{\mathbf{k}} C^\dagger_{\mathbf{k}\downarrow}|\downarrow\rangle \tag{6.304}$$

$$\phi_3 = \sum_{\mathbf{k}} d_{\mathbf{k}}[C^\dagger_{\mathbf{k}\uparrow}|\downarrow\rangle - C^\dagger_{\mathbf{k}\downarrow}|\uparrow\rangle] \tag{6.305}$$

$$\phi_4 = \sum_{\mathbf{k}} e_{\mathbf{k}}[C^\dagger_{\mathbf{k}\uparrow}|\downarrow\rangle + C^\dagger_{\mathbf{k}\downarrow}|\uparrow\rangle] \tag{6.306}$$

where $a_{\mathbf{k}}$, $b_{\mathbf{k}}$, $d_{\mathbf{k}}$, and $e_{\mathbf{k}}$ are coefficients and $|\uparrow\rangle$, $|\downarrow\rangle$ are configurations of the localized spins. Find the eigenvalue equations for each case, and identify the spin triplet and singlet states.

2. Evaluate the integral $\Lambda(\varepsilon)$ in the Kondo model for each $g(\varepsilon)$ and show that the coefficient of the term $\ln(\varepsilon/W)$ is proportional to $\rho(0)$. Also verify that the integral of $g(\varepsilon)$ over the entire band $(-W < \varepsilon < W)$ is unity.

$$\Lambda(\varepsilon) = \int_{-W}^{0} d\varepsilon' \frac{g(\varepsilon')}{\varepsilon' - \varepsilon} \tag{6.307}$$

$$\begin{aligned} &\text{(a)} \quad g(\varepsilon) = 1/(2W) \\ &\text{(b)} \quad g(\varepsilon) = 2\sqrt{W^2 - \varepsilon^2}/(\pi W^2) \\ &\text{(c)} \quad g(\varepsilon) = (W - |\varepsilon|)/W^2 \end{aligned} \tag{6.308}$$

3. Construct the zero-temperature collective state for the Kondo Hamiltonian when the local spin has $S = 1$. Find the states with $j = \frac{3}{2}$ and $j = \frac{1}{2}$. Which is preferred for $J < 0$?

4. Consider the collective state for the antiferromagnetic Kondo model when the bands are half-filled. Calculate $\Delta = -\delta E_a/W$ by solving (6.48) using the density of states in part (a) and (c) of Problem 2. Make a graph of Δ vs Λ defined in (6.52). Which case predicts the lowest binding energy?

5. Assume that an f level is split by the spin–orbit interaction into states with degeneracy 8 and 6, separated by an energy $\xi = 0.3$ eV. Derive an expression for $\delta = \varepsilon_f - \delta E$ assuming that $\delta \ll \xi$.

6. Derive the high-temperature ($\delta C \sim 1/T$) and low-temperature ($\delta C \sim T$) limits of the specific heat per impurity in the Anderson model.

7. Derive some vertex corrections to Eq. (6.113) by taking $n' = n \pm 1$. Can you draw a Feynman diagram for these processes?

8. Derive the energy current for the Anderson model. Then use it to derive the expression (6.122) for the Seebeck coefficient.

9. Calculate U and W for metallic aluminum. Estimate the bandwidth $W \approx E_F$. Calculate U using a charge density for the 3s atomic state in aluminum of $\rho_{3s}(r) = Nr^4 \exp(-\alpha r)$, where $\alpha = 2.8$a.u. What is the ratio of U/W?

10. How large a negative number can you make $v_c = v(\mathbf{q}) + U/2$ for the square lattice in two dimensions? Assume that the charge distribution on each atom is a Gaussian of the form $\phi(r) = \exp(-r^2/b^2)/(\pi b^2)^{3/2}$ and vary b to find the largest negative value of v_c at the corner of the Brillouin zone.

11. Evaluate the function $P^{(1)}(\mathbf{q}, 0)$ for the first-neighbor tight-binding model of the two-dimensional square lattice. Choose $\mathbf{q}$ to be the corner of the Brillouin zone $\mathbf{Q}a = \pi(\pm, \pm)$. Show that for this case that $\varepsilon(\mathbf{k} + \mathbf{Q}) = -\varepsilon(\mathbf{k})$ for all values of wave vector. Then show that $P^{(1)}(\mathbf{Q}, 0)$ is infinity at zero temperature when the band is exactly half full. In this case the Stoner criteria is satisfied for any value of U.

12. Use RPA in the extended Hubbard model to derive expressions for the susceptibilities $\chi_{\uparrow\downarrow}(q)$.

13. Find $P^{(1)}(q, \omega)$ for the jellium model of the free-electron gas, in the limits: (a) $x = 0$, and (b) x becomes large, where $x = \omega/(qv_F)$.

14. Evaluate the charge fluctuations in two dimensions using (6.155) in the limit of large x. For $d = 2$ the small wave vector limit is $v(\mathbf{q}) \to 2\pi e^2/(qA_0)$ where A_0 is the area of a unit cell.

15. For the pure Hubbard model ($v(q) = 0$):

    (a) Prove there are no RPA terms in the electron self-energy with two bubbles.
    (b) Start from the all-spin form in (6.175) and show that there is a term such as U^3P^2.
    (c) Find the exchange graphs which cancel the term in part (b), in order to agree with (a).

16. Derive the energy band dispersion $\varepsilon(\mathbf{k})$ for the hc lattice in the tight binding approximation. Include only first neighbor interactions, which are assumed to be isotropic. Find the reciprocal lattice vectors, and the shape of the Brillouin zone. What are the minimum and maximum values of $\varepsilon(\mathbf{k})$.

17. Derive the energy band dispersion for the pt lattice in the tight-binding approximation. Include only first neighbor interactions, which are assumed to be isotropic. Find the reciprocal lattice vectors, and the shape of the Brillouin zone. What are the minimum and maximum values of $\varepsilon(\mathbf{k})$.

18. Use the order parameter $A_{\sigma\bar{\sigma}}(\mathbf{Q})$ to solve for the AF phase of the sq lattice. In this case, the effective Hamiltonian is

$$H = \sum_{\mathbf{k}\sigma} [\varepsilon(\mathbf{k}) + Un_{\bar{\sigma}}]n_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma} \varepsilon(\mathbf{k}+\mathbf{Q}) + Un_{\bar{\sigma}}]n_{\mathbf{k}+\mathbf{Q}\sigma} - U\sum_{\mathbf{k}\sigma} A_{\sigma\bar{\sigma}}(\mathbf{Q})[C^{\dagger}_{\mathbf{k}+\mathbf{Q}\sigma}C_{\mathbf{k}\bar{\sigma}} + C^{\dagger}_{\mathbf{k}\sigma}C_{\mathbf{k}+\mathbf{Q}\bar{\sigma}}] \tag{6.309}$$

where $\mathbf{k}$ is summed over the AF Brillouin zone. One should get the same expression for the gap equation as found using the order parameter $A_{\mathbf{q}\sigma}$.

19. Derive a theory for the AF state in the hc lattice, where there is no $\mathbf{Q}$ vector since the Brillouin zone does not change size in going to the AF state. The first step is to choose a suitable order parameter.

20. Assume the density of one-particle states has the form $\rho(\varepsilon) = (2/\pi W^2)\sqrt{W^2 - \varepsilon^2}$. Use Bogoliubov theory to find the phase boundary between the PM and FM states.

21. Use the density of states of the prior problem to find the boundary of the AF phase using Stoner theory.

22. A single sheet of graphite has the honeycomb lattice which is bipartite. The sigma bonds are fully occupied by electrons. The π-bonds are half full which makes it a conductor. There is one electron per site when undoped. The density of states vanishes at half-filling. Assume it has the form:

$$\rho(E) = \frac{|E|}{D}, \qquad \text{for } -W \le E \le W \tag{6.310}$$

(a) Derive the relation between D and W assuming the band holds two electrons.
(b) Find a formula for the chemical potential as a function of filling $\mu(n)$.
(c) Derive the Stoner criteria for the onset of the FM state $U_c(n)$ in the Hubbard model.

23. Derive the Vilk–Tremblay sum rules for the extended Hubbard model.

24. Solve the Hubbard model approximately by taking

$$H_0 = -\mu\sum_{j\sigma} C^{\dagger}_{j\sigma}C_{j\sigma} + U\sum_{j} n_{j\uparrow}n_{j\downarrow}$$
$$V = -t\sum_{j\delta\sigma} C^{\dagger}_{j+\delta,\sigma}C_{j\sigma}$$

Calculate the Free energy to second order in V which are the terms $(\Omega_0, \Omega_1, \Omega_2)$.

(a) Use perturbation theory to evaluate the free energy of the Hubbard model to $O(U^2)$.
(b) Next show that the double site occupancy can be found from

$$\langle n_{j\uparrow}n_{j\downarrow}\rangle = \frac{1}{N}\frac{\partial\Omega}{\partial U}$$

(c) Use the result of (a) to show that the above average behaves as $(n/2)^2[1 - U/U_C]$ and estimate the critical value U_c.

Chapter 7

Electron–Phonon Interaction

7.1. FRÖHLICH HAMILTONIAN

The Fröhlich Hamiltonian describes the interaction between a single electron in a solid and LO (longitudinal optical) phonons:

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} C^{\dagger}_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \omega_0 \sum_{\mathbf{q}} a^{\dagger}_{\mathbf{q}} a_{\mathbf{q}} + \sum_{\mathbf{q}\mathbf{p}\sigma} \frac{M_0}{\sqrt{\nu}} \frac{1}{|q|} C^{\dagger}_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} (a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}})$$

$$M_0^2 = \frac{4\pi\alpha\hbar(\hbar\omega_0)^{3/2}}{\sqrt{2m}}, \qquad \varepsilon_p = \frac{p^2}{2m} \tag{7.1}$$

$$\alpha = \frac{e^2}{\hbar}\left(\frac{m}{2\hbar\omega_0}\right)^{1/2}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \tag{7.2}$$

This Hamiltonian was derived in Chapter 1, with the form of the interaction given in Sec. 1.3.5. The LO phonons are usually represented by an Einstein model, i.e., the phonon frequency $\omega_0 \equiv \omega_{\mathrm{LO}}$ is taken to be a constant. Since there is a single electron, the Hamiltonian may also be written as

$$H = \frac{p^2}{2m} + \omega_0 \sum_{\mathbf{q}} a^{\dagger}_{\mathbf{q}} a_{\mathbf{q}} + \sum_{\mathbf{q}} \frac{M_0}{\sqrt{\nu}} \frac{e^{i\mathbf{q}\cdot\mathbf{r}}}{|q|} (a_{\mathbf{q}} + a^{\dagger}_{-\mathbf{q}}) \tag{7.3}$$

where $\mathbf{r}$ and $\mathbf{p}$ are the conjugate coordinates of the electron. The unperturbed electron is taken to have free-particle motion with an effective mass m. Since there is only one electron in the problem, the results are independent of the statistics of the particle. The same results are obtained for any fermion or boson in the solid, such as holes, positions, etc., as long as they are free to move. The phonon modes are unaffected by the one electron in the solid, so the phonon self-energy is zero [actually of order $O(1/\nu)$], and the phonon Green's function $\mathscr{D}$ is always $\mathscr{D}^{(0)}$. The model also assumes that the motion is isotropic in direction and that the energy bands of the solid are nondegenerate. These rather restricted conditions describe what is called the *Fröhlich polaron problem*. The model actually applies in some cases to conduction bands of semiconductors and ionic solids which have their minimum at the Γ point and have an isotropic effective mass. For other symmetry points in the solid, one usually has to improve the model by taking into account anisotropy in the effective mass or degeneracy of the bands. The latter is important for holes in semiconductors. The Fröhlich polaron

problem was an important problem in mathematical physics during the 1950s. Numerous mathematical techniques were tried out on this problem. Several are described here: Brillouin–Wigner perturbation theory, Rayleigh–Schrödinger perturbation theory, strong coupling theory, and linked cluster theory. The Green's function method will be shown to be equivalent at zero temperature to Brillouin–Wigner theory. Several other methods were tried, including that of Low *et al.* (1953). The problem was most accurately solved by Feynman (1955). He introduced a variational method based on path integrals. After lengthy algebra, he obtained a result which even today is the best available. His theory is very lengthy and is not presented here. The Feynman results will, however, be used as the standard to which other theories are compared. The methods not discussed here, such as that of Low *et al.*, give poor results when compared with Feynman's method.

The polaron Hamiltonian (7.2) describes the motion of a particle while it is linearly coupled to a system of boson particles. In the Fröhlich Hamiltonian, the bosons are optical phonons in a polar solid. The classical picture has the particle exerting forces upon the ions, which respond and move. The ion motion creates new forces which act back upon the particle. The nonzero ion frequency makes the reaction forces, of the ions on the particle, retarded in time. The quantum nature of the phonons makes these forces occur in discrete units. In both the classical and quantum pictures, the ion motion is pictured as a polarization of the surrounding medium by the particle. The particle must drag this polarization with it during its motion through the solid, which affects its energy and effective mass.

7.1.1. Brillouin–Wigner Perturbation Theory

The Brillouin (1932, 1933)–Wigner (1935) perturbation theory method is a historical predecessor of the modern Green's functions method. It is equivalent to solving the equation for the energy spectrum E_p of a particle of momentum $\mathbf{p}$,

$$E_p = \varepsilon_p + \mathrm{Re}[\Sigma_{\mathrm{ret}}(p, E_p)] \tag{7.4}$$

where $\Sigma_{\mathrm{ret}}(p, E)$ is the retarded self-energy. The problem is isotropic, so the functions depend only upon the magnitude of the wave vector.

The method ignores the imaginary part of the self-energy of the electron. The particle properties are actually described by the spectral function:

$$A(p, E) = -2\,\mathrm{Im}\left[\frac{1}{E - \varepsilon_p - \Sigma_{\mathrm{ret}}(p, E)}\right] \tag{7.5}$$

If the imaginary part of the retarded self-energy is zero, it is replaced by an infinitesimal value $i\delta$, and the spectral function just becomes a delta function, whose argument is (7.4). The Brillouin–Wigner method is exact if the imaginary part of the retarded self-energy is zero. Then if the retarded self-energy function Σ_{ret} is found exactly, the exact result is obtained for the particle motion.

The self-energy function cannot be obtained exactly unless the Hamiltonian can be solved exactly. This result has not yet been achieved for (7.2). In practice, one usually evaluates a few terms in the perturbation expansion and thereby obtains an approximate $\Sigma_{\mathrm{ret}}(p, E)$. Equation (7.4) is solved with this approximate self-energy, which gives an approximate solution. This procedure is Brillouin–Wigner perturbation theory.

For one particle interacting with a set of optical phonons, the imaginary self-energy does vanish at zero temperature. This statement is only true for particles whose kinetic energy is

less than the phonon energies ω_0. For a particle of initial energy E_i to emit a phonon, it must go to a final state $E_f = E_i - \omega_0$. However, if E_i is less than ω_0, this equation cannot be satisfied since the energy E_f cannot be below the bottom of the band. The process is forbidden by energy conservation. Of course, a particle can absorb a phonon and increase its energy to $E_f = E_i + \omega_0$. This step is always possible as long as there are phonons in the system. But the number of phonons is proportional to the thermal occupation factor,

$$N_0 = \frac{1}{e^{\beta\omega_0} - 1} \tag{7.6}$$

which vanishes as $T \to 0$. At $T = 0$, the low-energy particle can neither absorb nor emit phonons. Since these are the only two loss mechanisms, the particle cannot lose or gain energy. Its mean-free-path is infinite, and the imaginary part of the retarded self-energy is zero. Brillouin–Wigner perturbation theory is exact at $T = 0$. Of course, it is exact only if the retarded self-energy is found exactly. In practice, that is impossible, and Brillouin–Wigner perturbation theory is usually a poor approximation.

The self-consistent energy function E_p is a smooth function of momentum at small values of p. It can be expanded in a power series:

$$E_p = E_0 + \frac{p^2}{2m^*} + O(p^4) \tag{7.7}$$

The quantity E_0 is the downward shift of the band minimum from polaron effects. E_0 is negative and gives the amount an electron with zero momentum lowers its energy by interacting with phonons. In Sec. 4.3 the self-energy of a fixed particle is $-\sum_{\mathbf{q}} M_{\mathbf{q}}^2/\omega_{\mathbf{q}}$, but it will be different now that the particle can move. The quantity m^* comes from the coefficient of the p^2 term in the momentum expansion. It is called the *effective mass*. In solids there are several effective masses. Each energy band will have its own curvature at the band minimum, which defines the effective band mass m_b. The band mass should be used for m in (7.2) and (7.3). The "b" subscript is omitted, but m is identical to m_b. The effective mass m^* in (7.7) is that resulting from the band curvature and polarons. It expresses the way that m (i.e., m_b) is changed by the polaron interactions. A formula for m/m^* was given previously in (3.160).

The first term in the perturbation series for the self-energy has one phonon in the self-energy diagram. The Feynman diagram is shown in Fig. 7.1(a). From the rules for constructing diagrams, the self-energy is

$$\Sigma^{(1)}(\mathbf{p}, ip) = \frac{M_0^2}{\beta\nu} \sum_{\mathbf{q}} \frac{1}{q^2} \sum_{iq_m} \mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip_n + iq_m)\mathscr{D}^{(0)}(\mathbf{q}, iq_m) \tag{7.8}$$

The Matsubara summation was given in Sec. 3.5:

$$\Sigma^{(1)}(\mathbf{p}, ip) = \frac{M_0^2}{\nu} \sum_{\mathbf{q}} \frac{1}{q^2} \left(\frac{N_0 + n_F}{ip_n + \omega_0 - \xi_{\mathbf{p}+\mathbf{q}}} + \frac{N_0 + 1 - n_F}{ip_n - \omega_0 - \xi_{\mathbf{p}+\mathbf{q}}} \right) \tag{7.9}$$

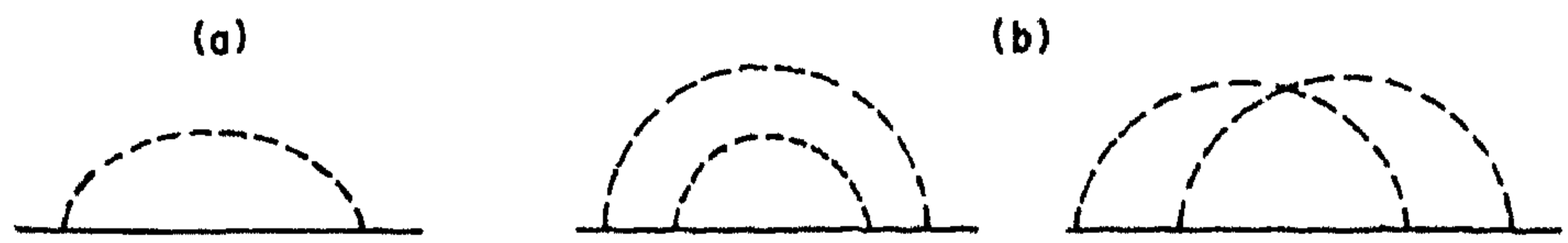


FIGURE 7.1

At zero temperature, $N_0 = 0$. The electron occupation factors n_F are all zero if there is only one particle in a band. Set $ip = E - \mu + i\delta$ so the real part of the retarded self-energy is ($\nu \to \infty$)

$$\mathrm{Re}[\Sigma^{(1)}_{\mathrm{ret}}(\mathbf{p}, E)] = \frac{M_0^2}{(2\pi)^3}\int \frac{d^3q}{q^2}\frac{1}{E - \omega_0 - (\mathbf{p}+\mathbf{q})^2/2m} \tag{7.10}$$

This result is also easy to obtain from the zero-temperature techniques of Chapter 2. The energy E is now measured from the bottom of the conduction band. A free particle has $E = 0$ when it has no kinetic energy. The wave vector integrals may be done exactly. Results are only given for $E < \omega_0$. The formulas at nonzero temperature, for all values of E, are given elsewhere (Mahan, 1972). Let $\nu = \cos(\theta) = \hat{p}\cdot\hat{q}$ be the angular variable. The three-dimensional integration is $d^3q = q^2 dq d\nu d\phi$ and the integral $\int d\phi = 2\pi$. The two remaining integrals are:

$$\mathrm{Re}[\Sigma^{(1)}_{\mathrm{ret}}(\mathbf{p}, E)] = -\frac{M_0^2}{(2\pi)^2}\int_0^\infty dq \int_{-1}^{1} d\nu \frac{1}{\omega_0 - E + (p^2 + q^2 + 2pq\nu)/2m}$$

The integrand remains unchanged if $q \to -q$, $\nu \to -\nu$. The answer is unchanged if the q limits are altered to $(-\infty, \infty)$ and the result is divided by 2; the two intervals $(-\infty, 0)$ and $(0, \infty)$ contribute equally. The integral can be simplified by changing the q variable to

$$x = \frac{q + p\nu}{\sqrt{2m}} \tag{7.11}$$

$$\mathrm{Re}[\Sigma^{(1)}_{\mathrm{ret}}(\mathbf{p}, E)] = -\frac{M_0^2\sqrt{2m}}{8\pi^2}\int_{-\infty}^{\infty} dx \int_{-1}^{1} d\nu \frac{1}{\omega_0 - E + \varepsilon_p(1-\nu^2) + x^2} \tag{7.12}$$

The x integral can be done first. The integral is of the form

$$\int_{-\infty}^{\infty} \frac{dx}{x^2 + A} = \frac{\pi\Theta(A)}{\sqrt{A}} \tag{7.13}$$

where A represents everything else in the denominator of (7.12). The integral is nonzero only when $A > 0$, which is always the case when $\omega_0 > E$. The remaining integral is just an arcsine:

$$\int_{-1}^{1} \frac{d\nu}{[\omega_0 - E + \varepsilon_p(1-\nu^2)]^{1/2}} = \frac{2}{\sqrt{\varepsilon_p}}\sin^{-1}\left(\frac{\varepsilon_p}{\omega_0 - E + \varepsilon_p}\right) \tag{7.14}$$

$$\mathrm{Re}[\Sigma^{(1)}_{\mathrm{ret}}(\mathbf{p}, E)] = -\frac{\alpha\omega_0^{3/2}}{\sqrt{\varepsilon_p}}\sin^{-1}\left(\frac{\varepsilon_p}{\omega_0 - E + \varepsilon_p}\right) \tag{7.15}$$

Equation (7.15) is the result for the one-phonon part of the retarded self-energy when $E > \omega_0$. The parameter E_0 is obtained from the self-energy at $p = 0$. By using the expansion for the arcsine, $\sin^{-1}(x) = x + x^3/6 + O(x^5)$, the $p \to 0$ limit is

$$\mathrm{Re}[\Sigma^{(1)}_{\mathrm{ret}}(0, E)] = -\frac{\alpha\omega_0^{3/2}}{\sqrt{\omega_0 - E_0}} \tag{7.16}$$

Equation (7.16) is the result for a zero-momentum particle. The result for a fixed particle is quite different. In Sec. 4.3 it is shown to be

$$\Delta E = -\frac{M_0^2}{\nu}\sum_{\mathbf{q}}\frac{1}{q^2} \tag{7.17}$$

The integral does not converge unless some cutoff is used at the high values of momentum. Of course, the cutoff could be the size of the Brillouin zone or alternately the spatial extent of the localized particle wave function. No cutoff is needed in the free-polaron problem because of the recoil energy of the electron. The zero-momentum result is different from the result for a fixed particle, a fact which is easily understood from the classical picture. The ions in the vicinity of the electron get either repulsed or attracted to the electron. Their motion changes the potential field felt by the electron. If the electron is fixed, it does not respond to this change. The ions seek their new equilibrium position without any change in the potential exerted upon them by the electron. A free electron can respond to the changing potential of the ions. When the ions move in response to the potential exerted upon them by the electron, the electron in turn begins to move in response to the changing potential of the ions. The final solution describes the coordinated motion of the electron and ions. For a free electron, the recoil of the electron makes polaron theory harder. A free polaron of zero momentum does not mean the electron is fixed. It is constantly moving, in response to the continual interplay with the force field of the ion. For $p = 0$, the motion has, on the average, zero momentum. But the electron is moving in a random, stochastic fashion. When it has nonzero momentum, it still has this random part of its motion, even as it is drifting. The total polaron momentum

$$\mathbf{P} = \mathbf{p} + \sum_{\mathbf{q}} \mathbf{q} a_{\mathbf{q}}^{\dagger} a_{\mathbf{q}} \tag{7.18}$$

commutes with the Hamiltonian and therefore is a constant of motion. Momentum is an acceptable eigenvalue to assign to the polaron.

The imaginary part of the one-phonon self-energy may also be obtained from (7.9). With $ip_n \to E - \mu$ and $n_F = 0$, the imaginary part is

$$-\mathrm{Im}[\Sigma_{\mathrm{ret}}^{(1)}(\mathbf{p}, E)] = \frac{\pi M_0^2}{(2\pi)^3} \int \frac{d^3q}{q^2} \{N_0 \delta(E + \omega_0 - \varepsilon_{\mathbf{q}+\mathbf{q}}) + (N_0 + 1)\delta(E - \omega_0 - \varepsilon_{\mathbf{p}+\mathbf{q}})\} \tag{7.19}$$

There are two terms in the brackets. The first term corresponds to the absorption of a phonon by the electron. This term is proportional to the phonon number density N_0, which vanishes at zero temperature. The second term comes from the emission of a phonon by the electron. It is proportional to the factor $N_0 + 1$, which is nonzero even at $T = 0$. However, this term is nonzero only if $E > \omega_0$, so that the particle has more energy than the phonon, and is zero when $E < \omega_0$. Therefore, $\mathrm{Im}[\Sigma_{\mathrm{ret}}^{(1)}(\mathbf{p}, E)] = 0$ at $T = 0$, unless $E > \omega_0$. The latter is possible in a high electric field.

The Tamm–Dancoff (TD) approximation constitutes solving Brillouin–Wigner perturbation theory with only the one-phonon self-energy:

$$E_p = \varepsilon_p + \mathrm{Re}[\Sigma_{\mathrm{ret}}^{(1)}(\mathbf{p}, E_p)] \tag{7.20}$$

For zero momentum, the TD approximation gives the particularly simple equation

$$E_0 = -\frac{\alpha \omega_0^{3/2}}{\sqrt{\omega_0 - E_0}} \tag{7.21}$$

The above is a cubic equation for E_0, which is easily solved. These results are summarized in Table 7.1 for several values of α. They are labeled E_{TD}.

The other columns are the Feynman results E_F and also the results of Rayleigh–Schrödinger perturbation theory $E_{RS}^{(1,2)}$. $E_{RS}^{(1)}$ is also evaluated in the one-phonon approximation

Table 7.1 Ground state energy of polaron in various theories: Feynman (E_F), Tamm–Dancoff (E_{TD}), and Rayleigh–Schrödinger ($E_{RS}^{(n)}$) for n phonons.

α	E_F/ω_0	E_{TD}/ω_0	$E_{RS}^{(1)}/\omega_0$	$E_{RS}^{(2)}/\omega_0$
1	−1.013	−0.76	−1.00	−1.016
2	−2.055	−1.31	−2.00	−2.064
3	−3.133	−1.80	−3.00	−3.143

to provide a fair comparison. This equation is very simple, $E_{RS}^{(1)} = -\alpha\omega_0$, as will be shown below. $E_{RS}^{(2)}$ also includes the two-phonon terms. The table shows that the Tamm–Dancoff approximation is a poorer approximation than Rayleigh–Schrödinger. It provides the larger error when compared with the Feynman result, particularly for values of $\alpha \sim 1$, which is called the intermediate coupling regime.

The Tamm–Dancoff gives poor results in the intermediate coupling regime. The poor answers occur because a gap was introduced in the excitation spectrum. The energy denominator in (7.10) has the difference between the initial energy of the particle E and the value in the intermediate state $\varepsilon_{\mathbf{p}+\mathbf{q}}$. However, E has the minimum value $E_0 < 0$, while $\varepsilon_{\mathbf{p}+\mathbf{q}}$ has the minimum value of zero. In order to get to the excited state, in this approximation, there is an excitation energy of E_0. Tamm–Dancoff is a poor approximation because there really is no excitation energy in the spectrum. The Tamm–Dancoff approximation just happens to insert a gap, which explains why it gives poor results for intermediate coupling.

The above results are suitable for a single particle in a band. A different approach must be used if there is a small but nonzero density n_0 of particles in the band. Then one must consider the change in the chemical potential $\delta\mu$. In insulators, the chemical potential μ has a negative value if the band minimum E_m is defined as zero energy. If n_0 is the density of particles, and λ is the deBroglie wavelength, then for the noninteracting system with Maxwell–Boltzmann statistics one has that $\mu - E_m = k_B T \ln(n_0\lambda^3)$. $\mu - E_m$ is negative if the argument of the logarithm is less than unity, which happens when n_0 is small. If polaron interactions cause a change δE_m in the energy of the band minimum, and if n_0 is unchanged, then the chemical potential must change according to $\delta\mu = \delta E_m$.

In the polaron problem, $E_m = 0$ and $\delta E_m = E_0$. How does $\delta\mu$ enter into the calculation? In going from (7.9) to (7.10), there was an analytic continuation $ip \to E - \mu - \delta\mu + i\delta$. Effectively E is replaced by $E - \delta\mu$ in all of the following equations. Since $\delta\mu = E_0$, (7.16) is replaced by

$$\mathrm{Re}[\Sigma_{\mathrm{ret}}^{(1)}(0, E_0)] = -\alpha\omega_0 \tag{7.22}$$

The right-hand side no longer depends upon E_0. The change in band minimum is $E_0 = -\alpha\omega_0$. This result now agrees with the Rayleigh–Schrödinger formula, which is given below. It is in much better agreement with the Feynman result than is the Tamm–Dancoff approximation. Brillouin–Wigner perturbation theory does better when more terms are included in the self-energy expression. The anticipated improvement is difficult to test since the next terms are quite formidable.

7.1.2. Rayleigh–Schrödinger Perturbation Theory

The Rayleigh–Schrödinger form of perturbation theory is the standard kind which is described in quantum mechanical textbooks such as Schiff (1955). It is also called *on the mass shell* perturbation theory. Energy and momentum are no longer separate variables. In

evaluating $\Sigma(p, E)$ the energy E is set equal to ε_p, so the self-energy is just a function of one variable p, or, equivalently, ε_p. Of course, if the imaginary part of the self-energy is actually zero, then energy and momentum are uniquely related. Rayleigh–Schrödinger is an exact procedure under the same conditions that Brillouin–Wigner perturbation theory is valid—a zero value for the imaginary self-energy. The discussion is confined to the case of zero temperature and $\varepsilon_p < \omega_0$ for which $\mathrm{Im}[\Sigma] = 0$.

The two perturbation theories have a direct link to the potential scattering theories of Sec. 4.1. The Rayleigh–Schrödinger perturbation theory is analogous to the reaction matrix theory of potential scattering, while Brillouin–Wigner is analogous to T-matrix theory. In the former, one calculates real quantities on the mass shells, i.e., with $E = \varepsilon_p$. In the latter, one calculates a complex quantity for a general value of E. Recall that the reaction matrix result was not equal to the real part of the T-matrix result. Similarly, the Rayleigh–Schrödinger electron–phonon self-energy of the electron is not found from the Brillouin–Wigner form by just setting $E = \varepsilon_p$ and taking the real part. This happens to work for the one-phonon self-energy but not when higher-order terms are included.

Successive terms in the perturbation theory for the self-energy function are called $\Sigma_{RS}^{(n)}(p)$. The superscript denotes the number of phonons in the self-energy term. The self-energy is real, and the signature "Re" before the self-energy is omitted. The one-phonon self-energy is

$$\Sigma_{\mathrm{RS}}^{(1)}(\mathbf{p}) = \frac{M_0^2}{(2\pi)^3}\int \frac{d^3q}{q^2}\frac{1}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p+q}} - \omega_0} \tag{7.23}$$

The energy denominator contains the difference between the initial state energy $\varepsilon_{\mathbf{p}}$ and the intermediate state energy $\varepsilon_{\mathbf{p+q}} + \omega_0$; the latter has one phonon excited. The summation over wave vectors is a summation over all intermediate states subject to momentum conservation. The self-energy is the same as the Brillouin–Wigner result (7.10) after replacing E by ε_p. This identity is true only for the one-phonon self-energies. For higher orders, in order to get Rayleigh–Schrödinger from the Brillouin–Wigner, replace E by ε_p *and* add some additional terms.

The wave vector integrals in (7.23) are elementary. Indeed, just take (7.15) and replace E by ε_p, with the result for $\varepsilon_p < \omega_0$

$$\Sigma_{RS}^{(1)}(\mathbf{p}) = -\frac{\alpha\omega_0^{3/2}}{\sqrt{\varepsilon_p}}\sin^{-1}\left(\frac{\varepsilon_p}{\omega_0}\right)^{1/2} \tag{7.24}$$

In the limit where $\varepsilon_p \to 0$, the simple result is:

$$\Sigma_{RS}^{(1)}(0) = -\alpha\omega_0 \tag{7.25}$$

It is no accident that the self-energy is just $-\alpha\omega_0$. The definition of the constant α was chosen to obtain a simple formula for the self-energy.

The polaron effective mass is derived by expanding the arcsine in a power series in $\sqrt{\varepsilon_p/\omega_0}$:

$$\Sigma_{RS}^{(1)}(\mathbf{p}) = -\alpha\omega_0\left[1 + \frac{1}{6}\frac{\varepsilon_p}{\omega_0} + O\left(\frac{\varepsilon_p}{\omega_0}\right)^2\right] \tag{7.26}$$

$$E_p = \varepsilon_p - \alpha\omega_0\left(1 + \frac{1}{6}\frac{\varepsilon_p}{\omega_0}\right) + O(p^4) \tag{7.27}$$

$$= -\alpha\omega_0 + \varepsilon_p\left(1 - \frac{\alpha}{6}\right) + O(p^4) \tag{7.28}$$

The effective mass is $m^* = m/(1 - \alpha/6)$. The polaron effects make the particle appear to be heavier than the band mass m. The extra mass arises because of the interaction between the electron and the ions. The electron causes a change in the equilibrium positions of the ions in its vicinity. The change in equilibrium position arises from the mutual interaction between the electron and ion, which was mentioned earlier. When the electron moves with a momentum $\mathbf{p}$, it must drag this ionic deformation with it. It takes energy to move this deformation. The drag is what causes the mass to increase as the polaron coupling constant is increased.

In the Tamm–Dancoff approximation, the formula (3.160) gives the effective mass $m^* = m(1 + \alpha/2)/(1 + \alpha/3)$. The details of this calculation are assigned as a problem. This result agrees with the Rayleigh–Schrödinger result at small values of α, since both formulas are proportional to $m^* = m[1 + \alpha/6 + O(\alpha^2)]$. However, they behave quite differently in the intermediate coupling regime. The Rayleigh–Schrödinger result predicts that something calamitous happens at $\alpha \approx 6$. A similar catastrophe is not implied by the Brillouin–Wigner formula, which is well behaved for all values of α. The Rayleigh–Schrödinger result is better. Something quite important does indeed happen at $\alpha \sim 6$: The particle becomes localized. This statement will be proved in the next subsection on strong coupling theory. It has already been remarked, in connection with Table 7.1, that the zero-momentum values are much better in the Rayleigh–Schrödinger picture.

At zero temperature, the two-phonon self-energy in the Rayleigh–Schrödinger theory is

$$\Sigma_{RS}(p) = \frac{M^4}{\nu^2} \sum_{\mathbf{q}_1\mathbf{q}_2} \frac{1}{q_1^2 q_2^2} \left(\frac{1}{\varepsilon_p - \varepsilon_{\mathbf{p}+\mathbf{q}_1} - \omega_0} \right) \left[\frac{1}{\varepsilon_p - \varepsilon_{\mathbf{p}+\mathbf{q}_1+\mathbf{q}_2} - 2\omega_0} \right.$$
$$\times \left(\frac{1}{\varepsilon_p - \varepsilon_{\mathbf{p}+\mathbf{q}_1} - \omega_0} + \frac{1}{\varepsilon_p - \varepsilon_{\mathbf{p}+\mathbf{q}_2} - \omega_0} \right)$$
$$\left. - \frac{1}{(\varepsilon_p - \varepsilon_{\mathbf{p}+\mathbf{q}_2} - \omega_0)^2} \right] \tag{7.29}$$

The wave vector integrals have been evaluated analytically for zero momentum. The integrals are complicated, with the result:

$$\Sigma_{RS}^{(2)}(0) = -\alpha^2\omega_0 \left[\ln\left(1 + \frac{3}{2\sqrt{2}}\right) - \frac{1}{\sqrt{2}} \right] = -0.1059\alpha^2\omega_0 \tag{7.30}$$

This result is combined with the earlier result for the one-phonon self-energy. The three-phonon self-energy was evaluated numerically by Smondyrev (1986), who computed the next terms in the Rayleigh–Schrödinger series for E_0 and m^*:

$$E_0 = -\omega_0[\alpha + 0.0159\alpha^2 + 0.000806\alpha^3 + O(\alpha^4)] \tag{7.31}$$

$$\frac{m}{m^*} = 1 + \frac{\alpha}{6} + 0.023627\alpha^2 + o(\alpha^3) \tag{7.32}$$

Some values are tabulated in Table 7.1 and agree well with the Feynman values. The exact ground state energy is a power series in α. Just the first three terms in this series are known. It is remarkable that the coefficients of the α^2 and α^3 terms are so small. Even for intermediate coupling values $1 < \alpha < 6$, the term $0.0159\alpha^2$ makes less than a 10% contribution to the total value of E_0. Another way to say this is that first-order perturbation theory is a good approximation even for intermediate coupling strengths. It is further evidence for the superiority of Rayleigh–Schrödinger over Brillouin–Wigner for the zero-temperature polaron

problem. It is assumed that higher-order terms also have a small coefficient and will also not contribute much toward E_0 for intermediate values of coupling constant.

It is natural to ask why the Rayleigh–Schrödinger method is so good. This point will be explained in detail in the next section. But it seems worthwhile to summarize some of these findings here. The independent boson model for a model system of electrons and phonons was solved in Chapter 4. The polaron model may not be solved exactly, but the Rayleigh–Schrödinger perturbation theory is related to the independent boson model. The self-energy for the interacting system of electrons and phonons arises from terms in which various numbers of phonons are virtually emitted into intermediate states. The one-phonon self-energy in the Rayleigh–Schrödinger picture corresponds to the assumption that all the phonons are virtually emitted independently of the others as in the independent-boson model. The two-phonon self-energy describes the correlations between the virtual emission of pairs of phonons. That is, $\Sigma_{RS}^{(1)}(p)$ contains the basic contribution from emitting one phonon, two phonons and all numbers of phonons. The next term $\Sigma_{RS}^{(2)}(p)$ describes the correlations between pairs of phonon emissions. The following term $\Sigma_{RS}^{(3)}(p)$ describes three-phonon correlations, which are not just pairwise correlations. The term $0.0159\alpha^2$ is small, apparently, because correlations are not important at intermediate coupling. For this reason, the $O(\alpha^3)$ terms are similarly small.

At first it seems surprising that $\Sigma_{RS}^{(1)}(p)$ has the self-energy from emissions of all different numbers of phonons but without correlations. However, it is similar to our findings for the independent boson model. There the exact self-energy was $-\sum_{\mathbf{q}} M_{\mathbf{q}}^2/\omega_{\mathbf{q}}$, yet this described a ground state which had a mixture of large numbers of phonons. If you take the Rayleigh–Schrödinger theory and set all kinetic energy terms equal to the same constant ε_0, you immediately recover the independent boson model in which the particle had constant energy. The one-phonon self-energy becomes

$$\Sigma_{RS}^{(1)}(p) = -\frac{M_0^2}{\nu}\sum_{\mathbf{q}}\frac{1}{q^2}\frac{1}{\omega_0} = -\sum_{\mathbf{q}}\frac{M_{\mathbf{q}}^2}{\omega_{\mathbf{q}}} \tag{7.33}$$

It is the exact self-energy when the kinetic energy is constant, which means that all higher self-energy terms are zero. An inspection of $\Sigma_{RS}^{(2)}(p)$ in (7.29) shows that it vanishes when all kinetic energies are a constant. However, it must vanish, since the exact self-energy for the independent boson model is just a linear function of α.

The independent boson model was solved in Sec. 4.3 by a variety of methods. It was not solved using Dyson's equation with all of its self-energy diagrams. At zero temperature, this procedure would correspond to trying to solve it by Brillouin–Wigner perturbation theory. In fact, the simple exact self-energy cannot be obtained this way—at least no one has ever succeeded in doing it. Apparently one has to evaluate all the terms in all orders of perturbation theory. It is not surprising that this approach does not work well for the polaron problem either.

A lover of antiques will enjoy that, for the polaron problem, the old-fashioned perturbation theory works better than Dyson's equations. One might ask whether this is a general feature. If so, why bother to learn about Green's functions? A general rule cannot be given. In some Hamiltonians the Rayleigh–Schrödinger method is best, and in others Dyson's equation is better. Coupled mode problems, such as solved for polaritons in Sec. 4.6, need to be described by Brillouin–Wigner perturbation theory. Another example: whenever the imaginary part of the self-energy is significant, it seems necessary to use a Green's function approach. The old-fashioned techniques do not allow for imaginary energies in a systematic

manner. The theory of strongly coupled superconductors, for example, could not be done without Green's functions. The actual message is that Green's functions are not the best way to solve all problems. Which problems are best solved by Green's functions? Unfortunately, at the moment, the only way to tell is by trial and error. No general rules are available which predict when one perturbation method is better than another. Of course, for the polaron problem, Brillouin–Wigner perturbation theory works fine as long as the chemical potential is renormalized along with changes in the band energies.

The word *polaron* describes the coupled system of electrons and ions. The self-energy is calculated for the mutually interacting system of electrons and phonons. Although it is the electron self-energy, part of the energy resides in the ions themselves. The ions are moved from their equilibrium position by the presence of the electron. The displacement takes some vibrational energy, which is part of the electron self-energy, because it follows the electron in its journey through the crystal. A simple analogy is the charge on a spring, which was discussed in Sec. 1.1. The application of the electric field causes a deformation of the spring. The final self-energy $-(eF)^2/2K$ contains a part which is the energy needed to compress the spring to the new equilibrium. In the polaron motion, the springs governing ion vibration occur locally wherever the polaron is at the moment.

The classical picture is that the ions in the polar lattice deform around the electron. The quantum picture is the same, except that the motion of the ions is quantized. The number of phonons is discrete. The amplitude of the ion displacements may not have a continuous range of values but only discrete amounts, which are phonons. If the average number of phonons around the electron is large, the quantization makes little difference. A laboratory spring does not appear to be quantized, although it surely is, but the displacements which usually are observed are so large, with such large quantum numbers, that the quantum nature is irrelevant. The same is true with our polaron if it has a large average number of phonons. These phonons are called the *phonon cloud.*

One may try to calculate the average number of phonons in the polaron. As with all things in polaron theory, it may only be calculated approximately. Since the Rayleigh–Schrödinger method is the best, it is used here. From first-order perturbation theory, the wave function of the electron at zero temperature is

$$\begin{aligned}\Psi_{\mathbf{p}}(\mathbf{r}) &= |\mathbf{p}\rangle + \sum_{\mathbf{p}'} \frac{|\mathbf{p}'\rangle\langle\mathbf{p}'|V|\mathbf{p}\rangle}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}'} - \omega_0} \\ &= \frac{1}{\sqrt{\nu}}\left[e^{i\mathbf{p}\cdot\mathbf{r}} + \frac{M_0}{\sqrt{\nu}}\sum_{\mathbf{q}}\frac{1}{q}\frac{e^{i(\mathbf{p}+\mathbf{q})\cdot\mathbf{r}}a_{\mathbf{q}}^{\dagger}}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} - \omega_0}\right]|0\rangle \end{aligned} \tag{7.34}$$

where $|0\rangle$ is the phonon vacuum, so $a_{\mathbf{q}}^{\dagger}|0\rangle$ has one phonon with wave vector $\mathbf{q}$. The total number of phonons is found by taking the expectation of the phonon number operator $\sum_{\mathbf{k}} a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}$ with this state:

$$\mathrm{Nu}(p) = \int d^3r\Psi_{\mathbf{p}}^{\dagger}(\mathbf{r})\sum_{\mathbf{k}} a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\Psi_{\mathbf{p}}(\mathbf{r}) \Big/ \int d^3r\Psi_{\mathbf{p}}\Psi_{\mathbf{p}} \tag{7.35}$$

The symbol Nu is used for number, to prevent confusion with other N symbols such as the thermal average number of phonons. The number operator on the first term in (7.34) is zero,

and on the second term it is unity. The first-order Rayleigh–Schrödinger prediction is

$$\mathrm{Nu}(p) = \frac{\Lambda}{1+\Lambda} \tag{7.36}$$

$$\Lambda(p) = \frac{M_0^2}{\nu}\sum_{\mathbf{q}}\frac{1}{q^2}\frac{1}{(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p+q}} - \omega_0)^2} \tag{7.37}$$

This wave vector integral is done by making the same variable changes that were used to obtain (7.12) [$\nu = \cos\theta, x = (q + p\nu)/\sqrt{2m}$]

$$\Lambda(p) = \frac{M_0^2\sqrt{2m}}{8\pi^2}\int_{-\infty}^{\infty} dx \int_{-1}^{1} d\nu \frac{1}{(\omega_0 - \nu^2\varepsilon_p + x^2)^2} \tag{7.38}$$

The two integrals give, in turn, the values for ($\omega_0 > \varepsilon_p$)

$$\int_{-\infty}^{\infty} dx \frac{1}{(\omega_0 - \nu^2\varepsilon_p + x^2)^2} = \frac{\pi}{2}\frac{1}{(\omega_0 - \nu^2\varepsilon_p)^{3/2}} \tag{7.39}$$

$$\int_{-1}^{1} d\nu \frac{1}{(\omega_0 - \nu^2\varepsilon_p)^{3/2}} = \frac{2}{\omega_0}\frac{1}{\sqrt{\omega_0 - \varepsilon_p}} \tag{7.40}$$

so the final result is

$$\Lambda(p) = \frac{\alpha}{2}\left(\frac{\omega_0}{\omega_0 - \varepsilon_p}\right)^{1/2} \Theta(\omega_0 - \varepsilon_p) \tag{7.41}$$

For a particle with zero momentum, the result is $\Lambda = \alpha/2$, $\mathrm{Nu} = \alpha/(2+\alpha)$. Nu(0) is always less than one, since the eigenstate only allowed for the creation of one phonon. It should be remembered that Nu(p) is an average number of phonons. The actual number must be an integer, which fluctuates from time to time about this average number.

These results apply for $\varepsilon_p < \omega_0$. An attempt to calculate the formula for $\varepsilon_p > \omega_0$ gives the result infinity. This answer is nonsensical, but the question was also nonsensical. A particle with kinetic energy larger than ω_0 will eventually emit a real phonon and lower its energy to $\varepsilon_p - \omega_0$. The number of phonons about the electron is not a stationary quantity when $\varepsilon_p > \omega_0$. It is not reasonable to try to evaluate a stationary matrix element such as Nu.

Another quantity to calculate is the mean-free-path of the electron. At zero temperature, it is infinity when $\varepsilon_p < \omega_0$. However, at nonzero temperatures, there is a nonzero probability N_0 that some thermally excited phonons do exist. One of them may be absorbed by the electron, thereby changing its energy and momentum. According to the golden rule (Schiff, 1955), the transition probability is

$$w = \frac{2\pi}{\hbar}\int \frac{d^3q}{(2\pi)^3}\frac{M_0^2}{q^2}N_0\delta(\varepsilon_{\mathbf{q}} + \omega_0 - \varepsilon_{\mathbf{p+q}}) \tag{7.42}$$

This formula is nearly the same as twice the imaginary part of the retarded self-energy given in (7.19). The main difference is the replacement of E by $\varepsilon_{\mathbf{p}}$. The second term of (7.19), proportional to $N_0 + 1$, has been omitted. It gives the rate of phonon emission; it is zero if $\varepsilon_p < \omega_0$. The integrals in the above expression are evaluated by changing variables to $dxd\nu$ to give

$$w = \frac{M_0^2\sqrt{2m}}{4\pi}N_0\int_{-\infty}^{\infty} dx \int_{-1}^{1} d\nu\delta(\omega_0 + \nu^2\varepsilon_p - x^2) \tag{7.43}$$

The two integrals may be evaluated in turn,

$$\int_{-\infty}^{\infty} dx\delta(\omega_0 + \nu^2\varepsilon_p - x^2) = \frac{1}{(\omega_0 + \nu^2\varepsilon_p)^{1/2}} \tag{7.44}$$

$$\int_{-1}^{1} d\nu \frac{1}{(\omega_0 + \nu^2\varepsilon_p)^{1/2}} = \frac{1}{\sqrt{\varepsilon_p}} \ln\left[\frac{\sqrt{\omega_0 + \varepsilon_p} + \sqrt{\varepsilon_p}}{\sqrt{\omega_0 + \varepsilon_p} - \sqrt{\varepsilon_p}}\right] \tag{7.45}$$

to give the scattering rate for phonon absorption. The scattering rate is the same as the inverse lifetime of the particle $w = 1/\tau$:

$$\frac{1}{\tau_p} = \alpha N_0 \sqrt{\frac{\omega_0^3}{\varepsilon_p}} \ln\left[\frac{\sqrt{\omega_0 + \varepsilon_p} + \sqrt{\varepsilon_p}}{\sqrt{\omega_0 + \varepsilon_p} - \sqrt{\varepsilon_p}}\right] \tag{7.46}$$

The mean-free-path l_p is found from the classical expression $l_p = \tau_p v_p$, so that one multiplies τ_p by p/m. At zero momentum, the lifetime is evaluated as

$$\frac{1}{\tau_0} = 2\alpha\omega_0 N_0 \tag{7.47}$$

This lifetime is used to calculate the electron mobility $\mu = e\tau_0/m^*$ as discussed in Chapter 8.

7.1.3. Strong Coupling Theory

The strong coupling theory for polarons was invented by Landau and Pekar (1946). Their theory was the first work on polarons, which even preceded the word *polaron*. Their theory, and its subsequent improvements, is now known to be valid at large values of α: hence the current name of strong coupling theory. The method of calculation is radically quite different from the prior perturbation theories. It is basically a variational calculation on a Gaussian wave function. The calculation is done first and the physics is discussed afterward.

The Hamiltonian (7.3) is rewritten so that the phonon operators appear as displacements $Q_{\mathbf{q}}$ and their conjugate momenta $P_{\mathbf{q}}$, which are chosen so they are dimensionless:

$$Q_{\mathbf{q}} = \frac{1}{\sqrt{2}}(a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}}), \qquad a_{\mathbf{q}} = \frac{1}{\sqrt{2}}(Q_{\mathbf{q}} + iP_{\mathbf{q}}) \tag{7.48}$$

$$P_{\mathbf{q}} = -\frac{i}{\sqrt{2}}(a_{\mathbf{q}} - a^\dagger_{-\mathbf{q}}), \qquad a^\dagger_{\mathbf{q}} = \frac{1}{\sqrt{2}}(Q_{\mathbf{q}} - iP_{\mathbf{q}}) \tag{7.49}$$

$$H = \frac{p^2}{2m} + \frac{\omega_0}{2}\sum_{\mathbf{q}}(P_{\mathbf{q}}^2 + Q_{\mathbf{q}}^2) + M_0\sqrt{\frac{2}{\nu}}\sum_{\mathbf{q}}\frac{Q_{\mathbf{q}}}{q}e^{i\mathbf{q}\cdot\mathbf{r}} \tag{7.50}$$

The wave function of the many-particle system $\Phi(\mathbf{r}; Q_{\mathbf{q}})$ must contain the coordinates of the electron $\mathbf{r}$ and the ion displacements $Q_{\mathbf{q}}$. We make what appears to be a drastic assumption: the electron is localized with a Gaussian wave function. Later this interpretation will soften, and it is shown that the particle is not quite localized. But for the moment assume that the total wave function is a simple product of electron and phonon coordinates:

$$\Phi(\mathbf{r}; Q_{\mathbf{q}}) = \phi(\mathbf{r})\psi_n(Q_{\mathbf{q}} + \delta Q_{\mathbf{q}}) \tag{7.51}$$

$$\phi(\mathbf{r}) = \left(\frac{\beta}{\sqrt{\pi}}\right)^{3/2} \exp\left(-\frac{\beta^2}{2}r^2\right) \tag{7.52}$$

where β is a variational parameter. The phonon wave functions ψ_n are the usual harmonic oscillator wave functions, except that they are centered about an equilibrium displacement $-\delta Q_{\mathbf{q}}$, which also needs to be determined. The first step in the calculation is to take the expectation of the Hamiltonian over the electron part of the coordinates. For the two r-dependent terms the Gaussian wave functions give

$$\mathscr{H}(Q_{\mathbf{q}}) = \int d^3r\phi(\mathbf{r})H\phi(\mathbf{r}) \tag{7.53}$$

$$\int d^3r\phi\frac{p^2}{2m}\phi = \frac{3\beta^2}{4m} \tag{7.54}$$

$$\int d^3r\phi(\mathbf{r})^2 e^{i\mathbf{q}\cdot\mathbf{r}} = e^{-q^2/4\beta^2} \tag{7.55}$$

The second integral is most easily evaluated in (x, y, z) coordinates. The expectation value of the Hamiltonian is

$$\mathscr{H}(Q_{\mathbf{q}}) = \frac{3\beta^2}{4m} + \frac{\omega_0}{2}\sum_{\mathbf{q}}(P_{\mathbf{q}}^2 + Q_{\mathbf{q}}^2) + \sum_{\mathbf{q}} L_{\mathbf{q}}Q_{\mathbf{q}} \tag{7.56}$$

$$L_{\mathbf{q}} = M_0\sqrt{\frac{2}{\nu}}\frac{e^{-q^2/4\beta^2}}{q} \tag{7.57}$$

The next step is to choose the equilibrium displacement $\delta Q_{\mathbf{q}}$ so that the term linear in $Q_{\mathbf{q}}$ is eliminated:

$$\delta Q_{\mathbf{q}} = \frac{L_{\mathbf{q}}}{\omega_0} \tag{7.58}$$

$$\mathscr{H}(Q_{\mathbf{q}}) = \frac{\omega_0}{2}\sum_{\mathbf{q}}[P_{\mathbf{q}}^2 + (Q_{\mathbf{q}} + \delta Q_{\mathbf{q}})^2] + \frac{3\beta^2}{4m} - \frac{1}{2\omega_0}\sum_{\mathbf{q}} L_{\mathbf{q}}^2 \tag{7.59}$$

$$= \frac{\omega_0}{2}\sum_{\mathbf{q}}[P_{\mathbf{q}}^2 + (Q_{\mathbf{q}} + \delta Q_{\mathbf{q}})^2] + E(\beta) \tag{7.60}$$

The first term on the right describes the harmonic vibrations of the phonons about their new equilibrium positions $-\delta Q_{\mathbf{q}}$. They have the harmonic wave functions ψ_n in (7.51) and the eigenvalues $(n + \frac{1}{2})\omega_0$. The second term $3\beta^2/4m$ is the kinetic energy of the electron in the Gaussian wave function. The last term is the potential energy of interaction between the phonons and electron. The parameter β is varied to give the lowest energy for these last two terms. First the potential term must be evaluated:

$$\frac{1}{2\omega_0}\sum_{\mathbf{q}} L_{\mathbf{q}}^2 = \frac{M_0^2}{\omega_0}\int\frac{d^3q}{(2\pi)^3}\frac{e^{-q^2/2\beta^2}}{q^2} = \alpha\left(\frac{\beta^2\omega_0}{m\pi}\right)^{1/2} \tag{7.61}$$

$$E(\beta) = \frac{3\beta^2}{4m} - \alpha\left(\frac{\beta^2\omega_0}{m\pi}\right)^{1/2} \tag{7.62}$$

Set $B = \beta^2/2m$, the last two terms are

$$E(B) = \tfrac{3}{2}B - \sqrt{B}\alpha\left(\frac{2\omega_0}{\pi}\right)^{1/2} \tag{7.63}$$

$$\frac{dE}{dB} = \tfrac{3}{2} - \frac{\alpha}{2\sqrt{B}}\left(\frac{2\omega_0}{\pi}\right)^{1/2} \tag{7.64}$$

The variation shows that the minimum value is $\sqrt{B_0} = (\alpha/3)\sqrt{2\omega_0/\pi}$ so that the minimum energy is

$$E(B_0) = -\frac{\alpha^2\omega_0}{3\pi} = -0.106\alpha^2\omega_0 \tag{7.65}$$

As usual with bound state variational calculations, the potential energy is twice the kinetic energy, in agreement with the virial theorem. This result was obtained by Landau and Pekar. The energy is proportional to α^2 rather than α.

The physics behind the variational theory is simple. The electron mass is much smaller than the ion mass, so the electron moves much faster. Its motion creates a charge density, and the ions respond to this average density. The small polaron occurs because of a strong feedback loop. The ions move to create a local potential which traps the electron in a bound state. The extent of the bound state depends upon the average motion of the electrons. The ions displace in response to the average motion of the electrons.

In strong coupling theory, it is assumed that α is very large and one tries to evaluate the energy as a power series in $O(1/\alpha)$. The philosophy is very similar to that for deducing the correlation energy of the homogeneous electron gas in Secs. 5.1 and 5.2. One perturbation expansion was developed for small r_s and another for high r_s. The same is done for the polaron. The Rayleigh–Schrödinger expansion is valid at small α, and the strong coupling is valid at large α. The interpolation between these two limits is remarkably easy. The best available result for the strong coupling limit (Miyake, 1976) is

$$\lim_{\alpha\to\infty} E_0(\alpha) = -\omega_0[0.1085\alpha^2 + 2.836 + O(1/\alpha^2)] \tag{7.66}$$

The coefficient of the α^2 term is 0.1085, which is remarkably close to the Landau–Pekar result of 0.1061. This agreement shows that the Gaussian assumption is very accurate.

Figure 7.2 shows a plot of the ground state energy for Rayleigh–Schrödinger perturbation theory as the solid curve labeled RS; also shown by the dashed line is the strong coupling theory using (7.66). The strong coupling theory appears to predict a lower energy state for small values of α. This value is deceptive, since terms of $O(1/\alpha^2)$ have omitted in the asymptotic series, and these obviously diverge at small α. The strong coupling theory results should be believed only for values of α above, say, about 5 or 6. The two curves almost touch for values of $\alpha \sim 5$. The value $\alpha \sim 5$ is believed to be the crossover region between the two theories. The correct theory is the Rayleigh–Schrödinger result for values up to $\alpha \leq 5$ and the strong coupling theory for $\alpha > 5$. Such a curve is precisely the behavior of the Feynman theory.

How large in space is the wave function of the small polaron? From the minimal value of B_0, the value of β_0 is

$$\beta_0 = \frac{\alpha}{3}\sqrt{\frac{4\omega_0 m}{\pi\hbar}} \tag{7.67}$$

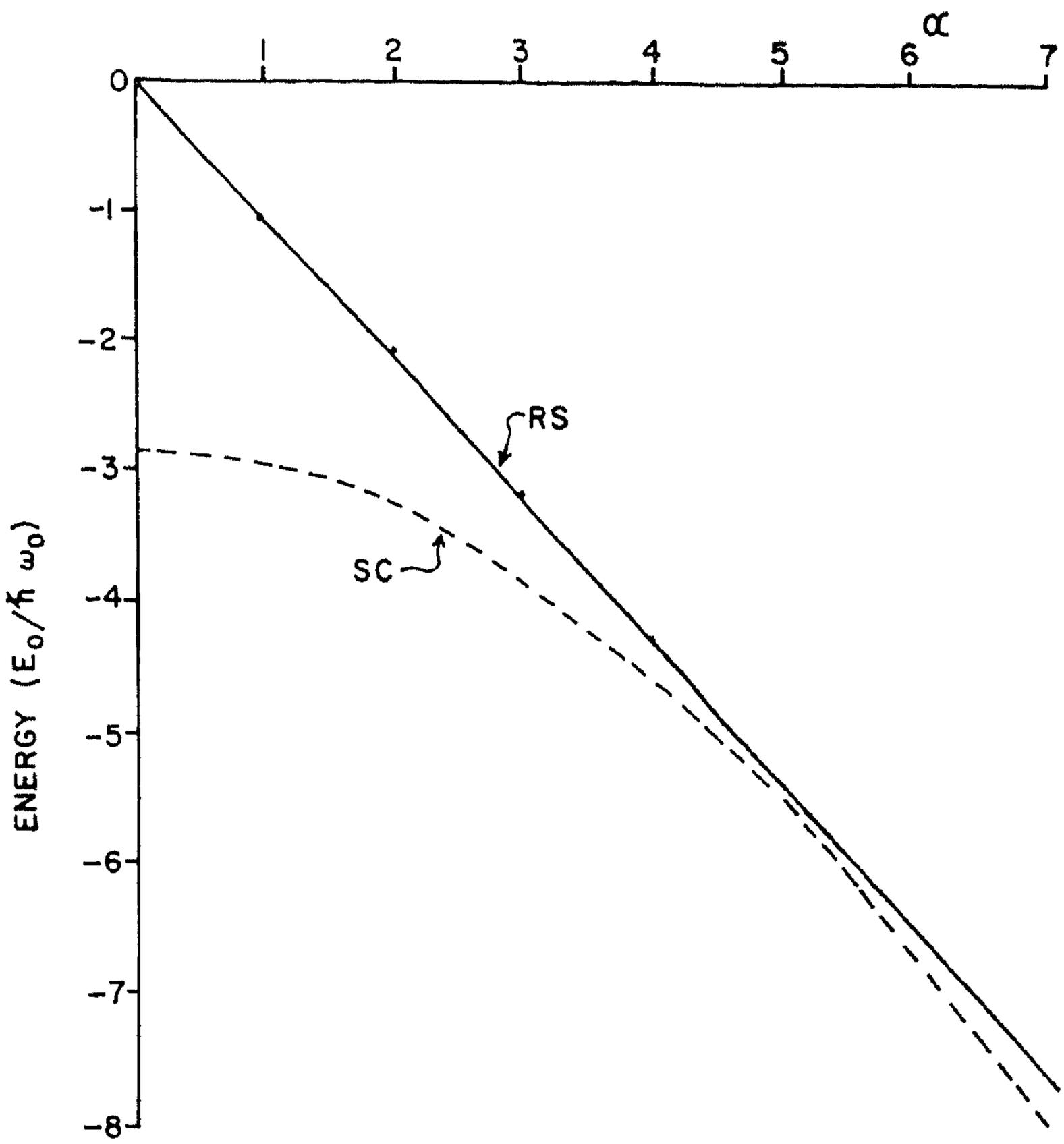


FIGURE 7.2 Energy of a polaron with zero momentum as a function of coupling constant. The solid line is the Rayleigh–Schrödinger theory, and the dashed line is the strong coupling theory.

β_0 has the dimensions of meters^{-1}, so its inverse is roughly the size of the localized wave function. For $\hbar\omega_0 \approx 0.03$ eV and m an electron mass, the polaron size is $\beta_0^{-1} \sim 40$ Å $/\alpha$. For α values around 5 to 6, the size is 7 Å, or about the size of the atomic unit cell. The localized electron will be influenced by the fact that the ions are atoms. The initial Hamiltonian (7.3) assumes a continuum theory for the ions, which is probably a reasonable approximation for small values of α but surely fails when the polaron size is of atomic dimension. Strong coupling theory cannot be applied to real solids without additional modifications to account for the atomic nature of the phonons. From now on polaron theory for $\alpha > 5$ must be considered a mathematical model with interesting properties. The useful piece of physics is that polarons become localized on atomic sites for values of α larger than 5 to 6. The Fröhlich Hamiltonian does not describe their behavior after they are localized.

The theory of polaron localization has a catch. One needs to know the band mass m to calculate α. Today one can deduce m from a good energy band calculation. In the early days they seldom had the required accuracy. Instead, it was customary to measure the polaron mobility and the effective mass, and the combined results will yield m, m^*, and α (see Hodby, 1972). The catch is that this procedure works only if the polaron is mobile. After it becomes localized, its mobility drops precipitously. The effective mass of a localized polaron cannot be measured by cyclotron resonance. It is difficult to deduce what the polaron constant α should have been after the particle is localized. It can only be deduced accurately if the particle is not localized. There are no experimental values of α above 4. Perhaps this value is the cutoff for localization of the polaron.

There are several other properties worth mentioning for polarons in strong coupling theory. The first is that the localization may occur anywhere. If the phonons were really a continuum, then the Gaussian packet would drift about and have an effective mass. The effective mass is quite large: Allcock (1956) has estimated it to be $m^* = 0.0208m\alpha^4$. For example, at $\alpha = 5$ it is $m^*/m = 13$. A similar estimate, $m^* = 0.0227m\alpha^4$, is given by Miyake (1976). The polaron is heavy because it has to drag with it the potential well of the phonons. The prediction from Rayleigh–Schrödinger perturbation theory is that the effective mass m^*/m was proportional to $1/(1 - \alpha/6)$ so that something important seemed to happen at $\alpha \sim 6$. What happens is localization, which makes the polaron much heavier. The heavy mass is not measurable because, as we said earlier, strongly coupled polarons are a model which we cannot apply to real systems without modification.

Another feature of strongly coupled polarons is that they have excited states which are also localized. For example, try to construct a polaron state which has p-wave symmetry. A suitable trial wave function might be

$$\phi(\mathbf{r}) = \left(\frac{2\beta^5}{\pi^{3/2}}\right)^{1/2} z e^{-\beta^2 r^2/2} \tag{7.68}$$

One could repeat the variational procedure with β as a variational parameter (see Problem 9 at the end of this chapter). It will be different from the s-wave value β_0, and it will be smaller. The wave function is more spread out, and the electron is less bound. It is called the *relaxed excited state* (see Kartheuser *et al.*, 1969). If the excited state were stable, one could observe actual optical transitions to this state from the ground state of s symmetry. The excited state is not stable, since it is degenerate in energy with the ground s state plus L phonons, where L is whatever number is necessary to make up the energy difference. Nevertheless, theoretical calculations show a sharp line in the optical spectra, which has been interpreted as this s- to p-wave transition. The strongly coupled polaron can create its own internal structure.

The strong coupling limit is calculated in the adiabatic approximation (see Allcock, 1962). The electron has sufficient binding energy that its oscillatory motion in the potential well is much faster than the vibrational frequency of the phonons. The phonons do not have time to adjust to the individual oscillations of the electron. Instead they adjust to the average motion of the electrons. The ions are treated as a rigid potential well, in which the electron adiabatically oscillates. A quite different picture applies to the weak coupling limit. There the phonon energy is larger than that of the electrons. The picture is that the phonons, or ion polarization, follow the electron during its motion.

7.1.4. Linked Cluster Theory

The linked cluster theory was advocated by Brout and Carruthers (1963) as a general method to attack many-body problems. It is surprising that the method waited so long to be applied to the Fröhlich polaron. It was not done until the work of Dunn (1975), although some earlier work of Mahan (1966) was similar. The surprise is even greater when it is realized that this method is an excellent way of obtaining the Green's function $G(p, t)$ for the polaron in the weak and intermediate coupling regimes. In the prior section, it was concluded that the Fröhlich polaron Hamiltonian is useful only for these coupling values. The linked cluster theory is applicable for all relevant coupling strength. One disagreeable aspect of the technique is that the final numbers have to be generated on the computer, since the formulas are

too complicated to permit analytical evaluation. This work is the price one pays for good Green's functions.

The linked cluster methods were described in Sec. 3.6. A problem with one electron may be worked in real time. The real-time formalism is valid even at nonzero temperatures. The Matsubara technique is not needed since the interactions do not change the thermodynamic averages. That is, when writing the Hamiltonian as $H = H_0 + V$, H_0 can be used in $\exp(-\beta H)$, rather than H, since V does not affect the thermodynamic averages. The only averaging is done over the phonons, and one electron in a macroscopic solid does not alter the phonon energies. This substitution, of using H_0 for H in the thermodynamic averaging, is not appropriate in a many-particle system where the particles influence the phonon modes—as in a metal or a heavily doped semiconductor. It is acceptable in the one-electron Fröhlich Hamiltonian. For the phonon Green's function at nonzero temperature, take

$$D^{(0)}(t) = -i[(N_0 + 1)e^{-i\omega_0|t|} + N_0 e^{i\omega_0|t|}] \tag{7.69}$$

$$N_0 = \frac{1}{e^{\beta\omega_0} - 1} \tag{7.70}$$

where $D^{(0)}(t)$ has no $\mathbf{q}$ dependence, since the phonons are assumed to have no dispersion. In this case $D^{(0)}(t) = D(t)$ since all the phonon self-energies are zero. The only concern is to evaluate the electron Green's function. By adapting the formulas of Sec. 3.6 to real time, the Green's function for the electron may be written as an exponential function of momentum and time:

$$G(\mathbf{p}, t) = G^{(0)}(t)\exp[F(\mathbf{p}, t)] \tag{7.71}$$

$$G^{(0)}(\mathbf{p}, t) = -i\Theta(t)e^{-\varepsilon_p t} \tag{7.72}$$

The function $F(\mathbf{p}, t)$ is generated as a series of terms, which are obtained by the following systematic procedure. The electron Green's function

$$G(\mathbf{p}, t) = -i\langle 0|Te^{iHt}C_{\mathbf{p}\sigma}e^{-iHt}C^\dagger_{\mathbf{p}\sigma}|0\rangle = -i\Theta(t)\langle T\hat{C}_{\mathbf{p}\sigma}(t)U(t)\hat{C}^\dagger_{\mathbf{p}\sigma}\rangle$$

is evaluated in the interaction representation:

$$\begin{aligned} G(\mathbf{p}, t) &= -i\Theta(t)\sum_n W_n(\mathbf{p}, t) \\ W_n(\mathbf{p}, t) &= \frac{(-i)^{2n}}{(2n)!}\int_0^t dt_1 \cdots \int_0^t dt_{2n}\langle T\hat{C}_{\mathbf{p}\sigma}(t)\hat{V}(t_1)\cdots\hat{V}(t_{2n})\hat{C}^\dagger_{\mathbf{p}\sigma}(0)\rangle \end{aligned} \tag{7.73}$$

The S-matrix expansion generates a series of scalar functions $W_n(\mathbf{p}, t)$. They are resummed as an exponential series of terms $F_n(\mathbf{p}, t)$

$$\begin{aligned} G(\mathbf{p}, t) &= -i\Theta(t)\sum_n W_n(\mathbf{p}, t) = -i\Theta(t)\exp\left[-i\varepsilon_p t + \sum_{n=1}^{\infty} F_n(\mathbf{p}, t)\right] \\ F(\mathbf{p}, t) &= \sum_{n=1}^{\infty} F_n(\mathbf{p}, t) \end{aligned} \tag{7.74}$$

These two series may be equated term by term by assuming that both W_n and F_n are proportional to powers of the coupling constant α^n. That is, define W_n to be all of the terms in the S-matrix expansion which are proportional to α^n. Similarly, when the series is resummed as an exponential, F_n contains all the terms proportional to α^n. For the polaron problem, the terms in the S-matrix are zero when l is odd, since they contain an odd number of phonon

creation or destruction operators. Only the terms with even l contribute, and $n = 2l$. With this convention, the relationship between some terms in the two series are:

$$F_1 = e^{i\varepsilon_p t} W_1 \tag{7.75}$$

$$F_2 = e^{i\varepsilon_p t} W_2 - \frac{1}{2!} F_1^2 \tag{7.76}$$

$$F_3 = e^{i\varepsilon_p t} W_3 - F_1 F_2 - \frac{1}{3!} F_1^3 \tag{7.77}$$

The terms are listed only up to $n = 3$, since that is already beyond what has been evaluated for the polaron problem. So far only the first two terms have been computed. However, these appear to be adequate to describe the polaron Green's function for low and intermediate coupling. The term W_1, or equivalently F_1, was first evaluated by Mahan (1966). This term has the form

$$W_1(\mathbf{p}, t) = \frac{(-i)^2}{2!} \int_0^t dt_1 \int_0^t dt_2 \langle T \hat{C}_{\mathbf{p}\sigma}(t) \hat{V}(t_1) \hat{V}(t_2) \hat{C}^\dagger_{\mathbf{p}\sigma}(0) \rangle \tag{7.78}$$

Write the Fröhlich Hamiltonian as $H = H_0 + V$,

$$H_0 = \sum_{\mathbf{p}\sigma} \varepsilon_p C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \omega_0 \sum_{\mathbf{q}} a^\dagger_{\mathbf{q}} a_{\mathbf{q}} \tag{7.79}$$

$$\hat{V}(t) = \frac{M_0}{\sqrt{\nu}} \sum_{\mathbf{q}\mathbf{p}\sigma} \frac{e^{-it(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}})}}{q} C_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} \hat{A}_{\mathbf{q}}(t) \tag{7.80}$$

$$\hat{A}_{\mathbf{q}}(t) = a_{\mathbf{q}} e^{-i\omega_{\mathbf{q}} t} + a^\dagger_{-\mathbf{q}} e^{-i\omega_{\mathbf{q}} t} \tag{7.81}$$

When evaluating the correlation function in (7.78), the phonon operators just give the phonon Green's function (7.69), and the electron correlation function can be written as a product of three $G^{(0)}$ by using Wick's theorem. The expression to evaluate is

$$\begin{aligned} W_1(\mathbf{p}, t) = &- \int \frac{d^3 q}{(2\pi)^3} \frac{M_0^2}{q^2} \int_0^t dt_1 \int_0^t dt_2 D(t_1 - t_2) \\ &\times G^{(0)}(\mathbf{p}, t - t_1) G^{(0)}(\mathbf{p} + \mathbf{q}, t_1 - t_2) G^{(0)}(\mathbf{p}, t_2) \end{aligned} \tag{7.82}$$

For one particle in a band, the electron Green's function is $G^{(0)}(\mathbf{p}, t) = -i\Theta(t) \exp[-i\varepsilon_p t]$. The above expression simplifies to

$$\begin{aligned} W_1(\mathbf{p}, t) = &- e^{-i\varepsilon_p t} \int \frac{d^3 q}{(2\pi)^3} \frac{M_0^2}{q^2} \int_0^t dt_1 \int_0^{t_1} dt_2 e^{i(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}})(t_1 - t_2)} \\ &\times [(N_0 + 1) e^{-i\omega_0 (t_1 - t_2)} + N_0 e^{i\omega_0 (t_1 - t_2)}] \end{aligned} \tag{7.83}$$

The two time integrals may be done without difficulty:

$$e^{i\varepsilon_p t} W_1(\mathbf{p}, t) = F_1(\mathbf{p}, t) \tag{7.84}$$

$$= -\int \frac{d^3q}{(2\pi)^3} \frac{M_0^2}{q^2} \left[it\left(\frac{N_0+1}{\Omega_-(\mathbf{p},\mathbf{q})} + \frac{N_0}{\Omega_+(\mathbf{p},\mathbf{q})} \right) + (N_0+1)\frac{1-e^{it\Omega_-}}{\Omega_-(\mathbf{p},\mathbf{q})^2} + N_0 \frac{1-e^{it\Omega_+}}{\Omega_+(\mathbf{p},\mathbf{q})^2} \right] \tag{7.85}$$

$$\Omega_\pm(\mathbf{p}, \mathbf{q}) = \varepsilon_\mathbf{p} - \varepsilon_{\mathbf{p}+\mathbf{q}} \pm \omega_0 \tag{7.86}$$

The function $F_1(\mathbf{p}, t)$ is determined. It has several features which are immediately interesting. There is a linear term in t. Its coefficient is just Rayleigh–Schrödinger one-phonon self-energy

$$\Sigma_{RS}^{(1)}(p) = \int \frac{d^3q}{(2\pi)^3} \frac{M_0^2}{q^2} \left(\frac{N_0+1}{\Omega_-(\mathbf{p},\mathbf{q})} + \frac{N_0}{\Omega_+(\mathbf{p},\mathbf{q})} \right) \tag{7.87}$$

It is just the real part of the self-energy, since the principal part is taken of the energy denominator. The self-energy is appropriate for nonzero temperatures. The result (7.23), for zero temperature, is obtained by setting $N_0 \to 0$. The first term describes processes whereby an electron of momentum $\mathbf{p}$ emits a phonon of wave vector $\mathbf{q}$ and goes to the state $\mathbf{p}+\mathbf{q}$. The second term describes processes where an electron in state $\mathbf{p}$ absorbs a phonon of $\mathbf{q}$ and goes to $\mathbf{p}+\mathbf{q}$. The energy denominator contains $\pm\omega_0$ depending on whether the intermediate state $\mathbf{p}+\mathbf{q}$ has one more or one less phonon than the initial state. The evaluation of these wave vector integrals is assigned as a problem.

The energy denominator $\Omega_+(\mathbf{p}, \mathbf{q})$ can vanish during the $\mathbf{q}$ integration. It enters one term in F_1 as $(\Omega_+)^{-2}$, which gives a possible divergence. However, there is no real divergence. Collect together all the terms in Ω_+, including those from the self-energy, and find the combination

$$\frac{1}{(\Omega_+)^2}\{1 + it\Omega_+ - e^{it\Omega_+}\} \tag{7.88}$$

which is well behaved when $\Omega_+ \to 0$. There is no actual divergence in the wave vector integration.

Most numerical work has been done for the momentum state $\mathbf{p} = 0$. In this case $F_1(0, t)$ can be obtained analytically in terms of the Fresnel integrals (Mahan, 1966) ($s = t\omega_0$)

$$E_2(x) = \int_0^x \frac{dte^{it}}{\sqrt{2\pi t}} \tag{7.89}$$

$$V_1(0, t) = -\frac{\alpha}{\sqrt{i\pi}} \left\{ (N_0+1)\left[-i\sqrt{s}e^{-is} + (2s+i)\sqrt{\frac{\pi}{2}}\, E_2(s)^* \right] + N_0\left[i\sqrt{s}e^{is} + (2s-i)\sqrt{\frac{\pi}{2}}\, E_2(s) \right] \right\} \tag{7.90}$$

The easiest way to derive this result is to follow the method of Dunn (1975). His suggestion is to go back to (7.83) and do the wave vector integrals before the time integrals t_1, and t_2

$$F_1(0, t) = -\int \frac{d^3q}{(2\pi)^3} \frac{M_0^2}{q^2} \int_0^t dt_1 \int_0^{t_1} dt_2 e^{-i\varepsilon_\mathbf{q}(t_1-t_2)}[(N_0+1)e^{-i\omega_0(t_1-t_2)} + N_0 e^{i\omega_0(t_1-t_2)}] \tag{7.91}$$

The wave vector integral has the form, for ($t > 0$)

$$\int_0^\infty dq e^{-itq^2/2m} = \sqrt{\frac{m\pi}{2it}} \tag{7.92}$$

Since the integrand is only a function of $t_1 - t_2$, change the integration variable to $t' = t_1 - t_2$, and then change the order of integrations to get

$$F_1(0, t) = -\frac{\sqrt{m}M_0^2}{2\pi^{3/2}} \int_0^t \frac{dt'}{\sqrt{it'}}(t - t')[(N_0 + 1)e^{-i\omega_0 t'} + N_0 e^{i\omega_0 t'}] \tag{7.93}$$

The time integrals are now recognized as just Fresnel integrals, or integrals related to them. The result (7.90) is derived easily. An approximate Green's function is obtained when we approximate $F(0, t)$ by its first term $F_1(0, t)$. The approximate expression can be numerically Fourier-transformed to obtain the spectral function

$$A^{(1)}(E) = -2\,\mathrm{Im}\left[(-i)\int_0^\infty dt e^{iEt} e^{F_1(0,t)}\right] \tag{7.94}$$

In this way, an approximation is derived for the imaginary part of the Green's function. The real part may also be obtained.

The accuracy of this approximate Green's function is not obvious. About the only way to judge is to also evaluate the next term $F_2(\mathbf{p}, t)$, which was done by Dunn. The next term $W_2(\mathbf{p}, t)$ produces three different terms, which correspond to the three Feynmann diagrams in Fig. 7.3. Usually such Feynman diagrams are associated with Dyson's equation, which is not the case here. Such diagrams are convenient to use, but here they have a different interpretation. For example, the diagram in Fig. 7.3(c) is not evaluated in Dyson's theory, since it is just the square of the one-phonon term. This identity is not true in the linked cluster theory, so it must now be evaluated. In fact the theory is not one of "linked clusters" at all, so the name is inappropriate. All diagrams are evaluated. A more descriptive name would be something like *exponential resummation*, which lacks pizzazz.

Some numerical results of Dunn (1975) are shown in Fig. 7.4. In Fig. 7.4(a) the spectral functions for $\mathbf{p} = 0$ and $k_B T = 0.4\omega_0$ are shown for three values of α. Dunn found that for $\alpha < 1$ the two-phonon results did not change the spectral function at all. The curves calculated with $F = F_1 + F_2$ were identical with Mahan's, which used $F = F_1$. For higher values of α there were changes, but they were not dramatic. Figure 7.4(b) shows a comparison of the one-phonon and two-phonon spectra functions for $\alpha = 6$. The theory is not applicable at this high a value, because one should be using strong coupling theory with its internal excited states. However, one can conclude that the linked cluster theory converges rapidly for low and intermediate coupling strengths.

The spectral functions shown in Fig. 7.4 describe the Green's functions for all frequencies ω. The low-energy exponential tail is for states below the conduction band minimum. The peaks at multiples of ω_0 are states with different numbers of phonons associated with the polaron. These are similar to those found for the independent boson model,

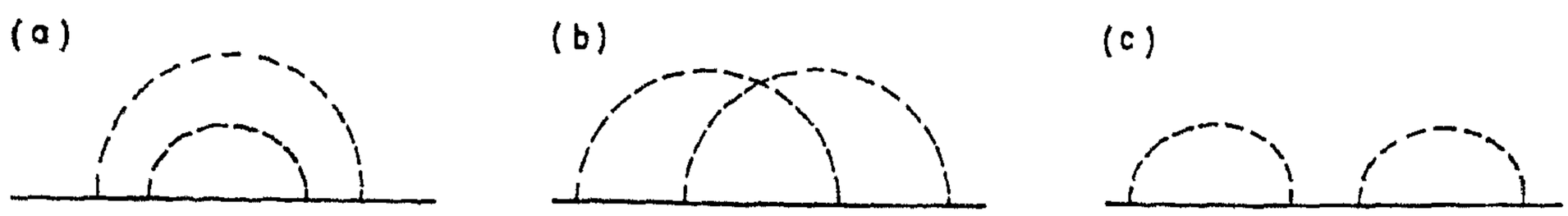


FIGURE 7.3

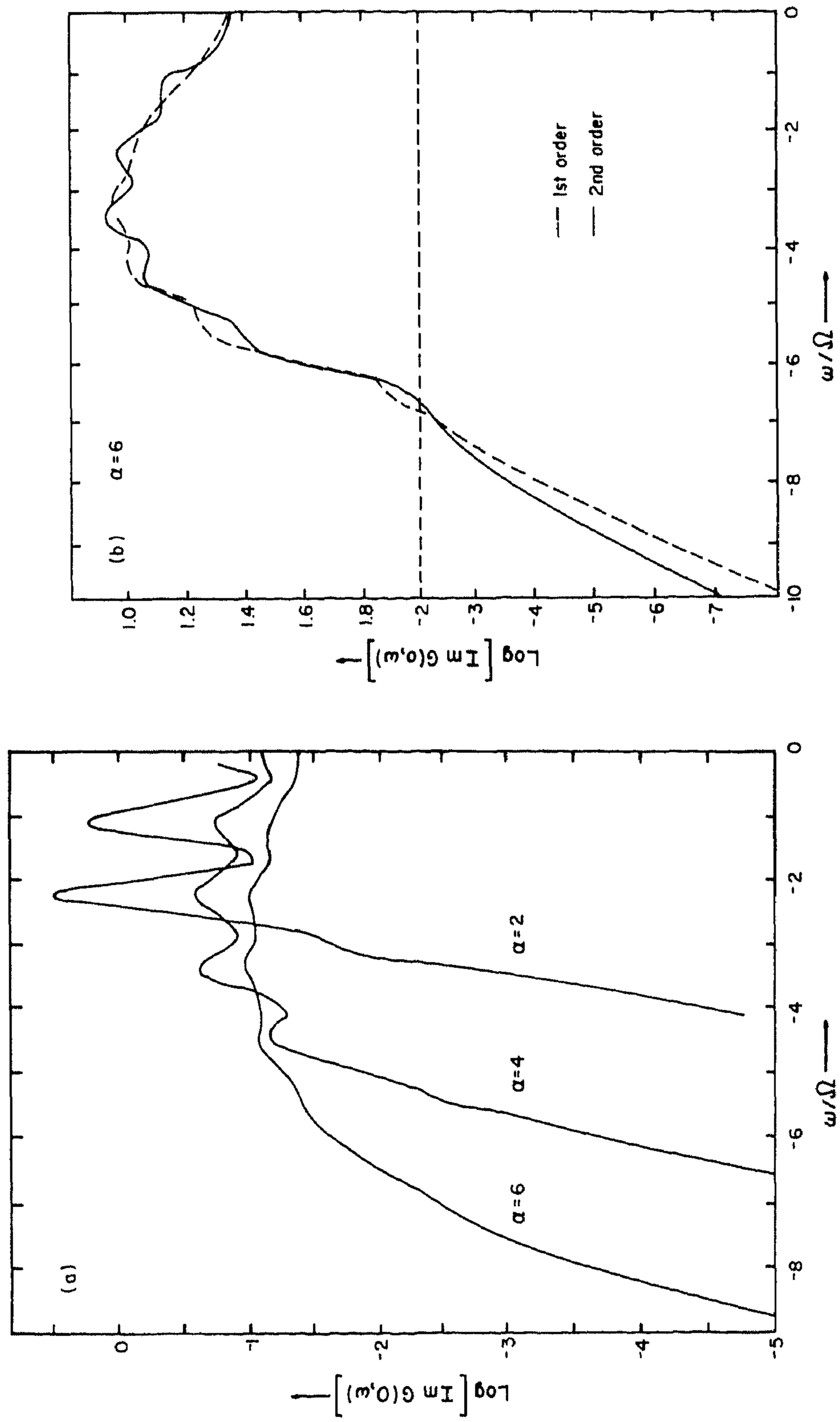


FIGURE 7.4 The spectral function of the polaron with zero momentum and $\beta\omega_0 = 2.5$. (a) Results for three coupling strengths, calculated with one- and two-phonon clusters. (b) A comparison of the results for one-phonon (dashed line) and one- plus two-phonon (solid line) clusters. *Source:* Dunn (1975) (used with permission).

except that here the peaks are broadened by the recoil of the electron. For $\alpha = 2$ the lowest two peaks are the main ones. But for $\alpha = 4$ and 6 the lowest peaks have small amplitude, which shows the influence of the renormalization factor.

The one-phonon term F_1 is very similar in form to the independent boson model. Compare (7.84) with (4.257) and (4.259). The only difference is the substitution of $\Omega_\pm = \varepsilon_{\mathbf{q}} - \varepsilon_{\mathbf{p}+\mathbf{Q}} \pm \omega_0$ in the independent boson model, instead of $\pm\omega_0$. F_1 is given the same interpretation. It describes the Green's function when successive emissions of phonons are independent. The approximate Green's function, with only F_1, does describe an electron coupled to many different numbers of phonons—but all uncorrelated. The addition of F_2 included correlations between pairs of phonon emissions. F_2 has little effect on either the self-energy or the spectral function itself at $\mathbf{p} = 0$.

The linked cluster theory is quite different from the Dyson's equation approach. There the one-phonon self-energy $\Sigma^{(1)}(p, E)$ adds only a one-phonon peak to the self-energy. The two-phonon peak comes from the two-phonon self-energy, etc. In order to get a full Green's function, with many phonon peaks, one needs to evaluate many self-energy diagrams. These higher-order diagrams are difficult to evaluate and have never been calculated. The Dyson's equation method has not been solved to the same accuracy as the linked cluster theory. Of course, one can also obtain $G(p, t)$ from Feynman theory.

7.2. SMALL POLARON THEORY

Polarons become "small" when they become localized, as in the strong coupling theory. In contrast, Fröhlich polarons are sometimes called "large." Small-polaron theory assumes that the size of the polaron corresponds with atomic dimensions. It recognizes the periodicity of the solid and thereby assumes that the motion of the particle is no longer translationally continuous. Instead, it assumes that the particle, usually an electron, may occupy an orbital state $\phi(\mathbf{r} - \mathbf{R}_j)$ centered on atomic site $\mathbf{R}_j$. The orbital states are identical on each site, so there is periodicity. The particle may move from site to site, exactly as in the tight-binding model. The motion from site to site may be caused by the overlap, or nonorthogonality, of the orbitals on adjacent sites. The phonons are coupled to the particle at whichever site it is on. The following Hamiltonian describes the motion of small polarons:

$$H = J \sum_{j,\delta} C_{j+\delta}^\dagger C_j + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} + V_{ep} \tag{7.95}$$

$$Q_j = \sum_{\mathbf{q}} X_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{R}_j}(a_{\mathbf{q}} + a_{-\mathbf{q}}^\dagger), \qquad X_{\mathbf{q}} = \sqrt{\frac{\hbar}{2MN\omega_{\mathbf{q}}}} \tag{7.96}$$

$$\mathbf{Q}_j = \sum_{\mathbf{q}} X_{\mathbf{q}} \hat{\xi}_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{R}_j}(a_{\mathbf{q}} + a_{-\mathbf{q}}^\dagger) \tag{7.97}$$

$$V_{ep}^{(1)} = D_1 \sum_{j\mathbf{q}} C_j^\dagger C_j Q_j \tag{7.98}$$

$$V_{ep}^{(2)} = D_2 \sum_j C_j^\dagger C_j \sum_\delta \hat{\delta} \cdot (\mathbf{Q}_j - \mathbf{Q}_{j+\delta}) \tag{7.99}$$

$$V_{ep}^{(3)} = D_3 \sum_{j\delta} C_{j+\delta}^\dagger C_j \hat{\delta} \cdot (\mathbf{Q}_j - \mathbf{Q}_{j+\delta}) \tag{7.100}$$

where C_j is the destruction operator for a particle on site $\mathbf{R}_j$. The summation δ is over the nearest neighbors. Spin indices are not important in this problem and are omitted. If (7.95) is compared with the Fröhlich Hamiltonian (7.2), only the first term is different. The present Hamiltonian (7.95), which includes the periodicity of the solid through the tight-binding model, is much more realistic.

Three different forms are provided for the electron–phonon coupling V_{ep}. The first one $V_{ep}^{(1)}$ treats the ion displacement Q_i as a scalar. This model is appropriate for organic solids, where the vibrational modes are those of the molecule. As an electron hops from molecule to molecule, it can excite vibrational modes of the molecule it is resting on. Then Q_i is the amplitude of the normal mode. The second form of the interaction $V_{ep}^{(2)}$ arises from the Coulomb interaction between the electron and the neighboring ions. When an electron is on ion $\mathbf{R}_j$ then there is a contraction or expansion of the separation between this ion and its neighbors. The third form of the electron–electron interaction $V_{ep}^{(3)}$ is due to the phonon modulation of the hopping. The first term in (7.95) has a hopping parameter $J(\delta)$ which must depends upon the distance to the neighbors. If this distance is modulated by ion vibrations, there is a corresponding electron–phonon interaction. In this case the coupling constant $D_3 = dJ/dr$. Both interactions $V^{(2)}$ and $V^{(3)}$ are present in all tight-binding solids. The first is appropriate for molecular solids.

The discussion of small polarons will follow the classic treatment of Holstein (1959), which followed the pioneering work of Tiablikov (1952) and Yamashita and Kurosawa (1958). They took $V_{ep}^{(1)}$ as the electron–phonon interaction. The same analysis works for $V_{ep}^{(2)}$.

The small-polaron Hamiltonian exhibits two types of behavior which are very different. Each type of behavior is described in detail. Afterwards, the interesting problem is discussed of the transition region between these limiting cases.

7.2.1. Large Polarons

The first class of behavior is "large" polaron motion, of the Fröhlich type. It occurs whenever the bandwidth zJ is large, where z is the coordination number. The condition on the bandwidth is stated later with more precision. When the bandwidth is large, the Hamiltonian is solved in wave vector space. Transform to collective coordinates:

$$C_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_j C_j e^{i\mathbf{k}\cdot\mathbf{R}_j} \tag{7.101}$$

$$H = zJ \sum_{\mathbf{k}} \gamma_{\mathbf{k}} C_{\mathbf{k}}^{\dagger} C_{\mathbf{k}} + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^{\dagger} a_{\mathbf{q}} + \sum_{\mathbf{kq}} C_{\mathbf{k}+\mathbf{q}}^{\dagger} C_{\mathbf{k}} M(\mathbf{k}, \mathbf{q})(a_{\mathbf{q}} + a_{-\mathbf{q}}^{\dagger})$$

$$\gamma_{\mathbf{k}} = \frac{1}{z} \sum_{\delta} e^{i\mathbf{k}\cdot\delta} \tag{7.102}$$

$$M(\mathbf{k}, \mathbf{q}) = X_{\mathbf{q}}\{D_1 - D_2 \hat{\xi}_{\mathbf{q}} \cdot \mathbf{F}(\mathbf{q}) + D_3 \hat{\xi}_{\mathbf{q}} \cdot [\mathbf{F}^*(\mathbf{k}+\mathbf{q}) - \mathbf{F}^*(\mathbf{k})]\} \tag{7.103}$$

$$\mathbf{F}(\mathbf{k}) = \sum_{\delta} \delta e^{i\mathbf{k}\cdot\delta} = -i\nabla\gamma_{\mathbf{k}} \tag{7.104}$$

where $\xi_{\mathbf{q}}$ is the unit vector for phonon polarization. The electron–phonon interaction has the usual form, except that the matrix element $M(\mathbf{k}, \mathbf{q})$ depends upon $\mathbf{k}$ as well as on $\mathbf{q}$. The dependence upon $\mathbf{k}$ comes from the phonon-modulated hopping. This term will be dropped ($D_3 = 0$), so that the electron–phonon matrix element $M(\mathbf{q}) \equiv M_{\mathbf{q}}$ will depend only upon $\mathbf{q}$.

The small-polaron Hamiltonian has a close resemblance to the Fröhlich polaron Hamiltonian. The only difference is the replacement of the free-particle energy $\varepsilon_k = k^2/2m$ by the tight-binding form $\varepsilon_{\mathbf{k}} = zJ\gamma_{\mathbf{k}}$, and $M_{\mathbf{q}}$ has a different dependence upon $\mathbf{q}$.

The wave vector summation for particles extends only over the Brillouin zone. In a realistic model of a solid, there would be many bands to be summed. For a large bandwidth, the particle will confine its motion to states near the bottom of the band. For J negative, the bottom of the band occurs at $\mathbf{k} = 0$. Expand $\gamma_{\mathbf{k}}$ about the point $\mathbf{k} = 0$, and find in cubic crystals $\varepsilon_{\mathbf{k}} = zJ(1 - (k\delta)^2/6)$, so that the particle has an effective mass of $m^{-1} = z|J|\delta^2/3$. If the polaron coupling strength $M_{\mathbf{q}}$ is small, this Hamiltonian can be described by weak coupling theory. There is a slight change in the particle's energy and effective mass because of polaron effects. These changes must correspond, in the tight-binding model, to a change in bandwidth. The polaron self-energy, in first-order Rayleigh–Schrödinger perturbation theory, is

$$\Sigma_{RS}^{(1)}(\mathbf{k}) = \sum_{\mathbf{q}} M_{\mathbf{q}}^2 \left\{ \frac{N_{\mathbf{q}} + 1 - n_F(\varepsilon_{\mathbf{k}+\mathbf{q}})}{\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}} - \omega_{\mathbf{q}}} + \frac{N_{\mathbf{q}} n_F(\varepsilon_{\mathbf{k}+\mathbf{q}})}{\varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}+\mathbf{q}} + \omega_{\mathbf{q}}} \right] \tag{7.105}$$

The wave vector integrals are rather hard to evaluate for most forms of $M_{\mathbf{q}}^2$, because the energy denominators have the inconvenient expressions $\varepsilon_{\mathbf{k}+\mathbf{q}} = zJ\gamma_{\mathbf{k}+\mathbf{q}}$. Numerical results have been obtained in one dimension (Brown *et al.*, 1997). Fröhlich-type polaron effects are obtained when the bandwidth is large and the polaron effects are small.

7.2.2. Small Polarons

Next consider the other limiting case of the small-polaron Hamiltonian (7.95). The polaron effects are assumed to be dominant and the bandwidth is small. The physical picture is that the polaron effects localize the particle on a site and that hopping occurs infrequently from site to site. The tight-binding term is the perturbation, while the particle–phonon term is large. The Hamiltonian is solved in position space for the particles without resorting to collective coordinates. The first step is to apply a canonical transformation which diagonalizes the last two terms in the Hamiltonian. These last two terms are the same as found in the exactly solvable models of Sec. 4.3 for the many-boson model. The canonical transformation has the same form:

$$\bar{H} = e^S H e^{-S} \tag{7.106}$$

$$S = -\sum_{j\mathbf{q}} n_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \frac{M(\mathbf{q})}{\omega_{\mathbf{q}}} (a_{\mathbf{q}} - a_{-\mathbf{q}}^\dagger) \tag{7.107}$$

$$\bar{H} = J \sum_{j\delta} C_{j+\delta}^\dagger C_j X_{j+\delta}^\dagger X_j + \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} - \Delta \sum_j n_j \tag{7.108}$$

$$\Delta = \sum_{\mathbf{q}} \frac{M(\mathbf{q})^2}{\omega_{\mathbf{q}}} \tag{7.109}$$

$$X_j = \exp\left[\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{R}_j} \frac{M(\mathbf{q})}{\omega_{\mathbf{q}}} (a_{\mathbf{q}} - a_{-\mathbf{q}}^\dagger) \right] \tag{7.110}$$

The polaron self-energy is Δ. The factors X_j were encountered previously in Sec. 4.3.4. They arise from the canonical transformation of the particle operators $e^S C_j e^{-S} = C_j X_j$. The number

operator $n_j = C_j^\dagger C_j$ commutes with S and is unaffected by the transformation. But the tight-binding term $JC_{j+\delta}^\dagger C_j$ produces factor $X_{j+\delta}^\dagger X_j$. The first term is not solvable exactly, so the canonical transformation did not diagonalize this Hamiltonian. However, so far no approximations have been made. The eigenstates and energy levels of $\bar{H}$ are the same as H. It is fruitful to investigate the solutions of $\bar{H}$. Use the interaction representation to set $\bar{H} = H_0 + V$:

$$H_0 = \sum_{\mathbf{q}} \omega_{\mathbf{q}} a_{\mathbf{q}}^\dagger a_{\mathbf{q}} - \Delta \sum_j n_j \tag{7.111}$$

$$V = J \sum_{j\boldsymbol{\delta}} C_{j+\delta}^\dagger C_j X_{j+\delta}^\dagger X_j \tag{7.112}$$

The exponents in the factor $X_{j+\delta}^\dagger X_j$ can be combined since they commute. Also assume that $M_{-\mathbf{q}}^* = M_{\mathbf{q}}$.

$$X_{j+\delta}^\dagger X_j = \exp\left[\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{R}_j}(1 - e^{i\mathbf{q}\cdot\boldsymbol{\delta}})\frac{M(\mathbf{q})}{\omega_{\mathbf{q}}}(a_{\mathbf{q}} - a_{-\mathbf{q}}^\dagger)\right] \tag{7.113}$$

Combining the two factors simplifies the form of this operator for taking the expectation of phonon operators. The perturbation V describes the hopping of the polaron from one site $\mathbf{R}_j$ to the neighboring site $\mathbf{R}_j + \boldsymbol{\delta}$. The amplitude of this process is

$$J\langle f|X_{j+\delta}^\dagger X_j|i\rangle \tag{7.114}$$

where $|i\rangle$ and $|f\rangle$ describe the phonon occupation numbers in the initial and final states of the transition. Since the factor $X_{j+\delta}^\dagger X_j$ permits phonons to be made or destroyed, the states $|i\rangle$ and $|f\rangle$ may have different numbers of phonons.

Holstein (1959) made a distinction between events in which the number of phonons were changed during the hop and events where they did not change. If the number of phonons is changed in the hop, it is an inelastic scattering process. The particle loses its phase coherence by this emission or absorption of phonons. Each hop becomes a statistically independent event. The particle motion is diffusive, since it has Brownian motion by randomly hopping from site to site. These events are called *nondiagonal transitions*.

The other situation is a *diagonal transition* in which all the phonon occupation numbers remain the same during the hop. The number of phonons $n_{\mathbf{q}}$ in each state $\mathbf{q}$ remains unaltered. (Note: $n_{\mathbf{q}}$ is the integer number of phonons, while $N_{\mathbf{q}}$ is the thermal average of $n_{\mathbf{q}}$.) This condition is rather strict. For example, if the number of phonons is unchanged but one state gains a phonon while another loses one, it is a nondiagonal transition. The phase coherence of a particle is maintained during a diagonal transition. If it is the most likely hopping event, the particle will hop from site to site while retaining phase coherence. Then it is a Bloch particle and forms energy bands. The diagonal part of the hopping probability is defined as

$$e^{-S_T} = \langle i|X_{j+\delta}^\dagger X_j|i\rangle \tag{7.115}$$

If the nondiagonal transition probability is small, the particle will form energy bands with a band energy

$$\bar{\varepsilon}_{\mathbf{k}} = zJ\gamma_{\mathbf{k}} e^{-S_T} \tag{7.116}$$

The effect of the polarons is to reduce the bandwidth by a factor $\exp(-S_T)$. The effective mass is increased by the same factor. In the Holstein picture, the diagonal hops contribute to

the real part of the self-energy and the nondiagonal transitions contribute to the imaginary part of the self-energy.

The large-polaron behavior, of the Fröhlich type, occurs when the diagonal transitions dominate. The small-polaron behavior, with diffusive hopping, occurs whenever the nondiagonal transitions dominate.

7.2.3. Diagonal Transitions

The diagonal and nondiagonal transition rates are obtained by using the same type of operator techniques introduced in Sec. 4.3. The diagonal transition rate is calculated from (6.2.5). The factor $\langle f|X_{j+\delta}^{\dagger}X_j|i\rangle$ has two operators in the exponential which are Feynman disentangled, with the destruction operators to the right. Using the techniques introduced into Sec. 4.3 gives

$$\lambda = e^{i\mathbf{q}\cdot\mathbf{R}_j}(1 - e^{i\mathbf{q}\cdot\boldsymbol{\delta}})\frac{M(\mathbf{q})}{\omega_{\mathbf{q}}} \tag{7.117}$$

$$e^{-S_T} = \Pi_{\mathbf{q}}\langle i|e^{\lambda a - \lambda^* a^\dagger}|i\rangle \tag{7.118}$$

The result is given by events with the same number of phonons in each state $\mathbf{q}$ for both $\langle i|$ and $|i\rangle$. It is assumed that the number of phonons in each state is determined by a thermal average. The summation is taken over each possible number $n_{\mathbf{q}}$ of phonons in each state, with the thermal probability $\exp(-\beta\omega_{\mathbf{q}}n_{\mathbf{q}})$ for this occurrence [see (4.239)]:

$$\begin{aligned} e^{-S_T} &= \Pi_{\mathbf{q}}\exp\left(\frac{|\lambda_{\mathbf{q}}|^2}{2}\right)\left[1 - e^{-\beta\omega_{\mathbf{q}}}\right]\sum_{n=0}^{\infty} e^{-\beta\omega_{\mathbf{q}}n}\langle n|e^{\lambda^* a^\dagger}e^{-\lambda a}|n\rangle \\ &= \Pi_{\mathbf{q}}\exp[-|\lambda_{\mathbf{q}}|^2(N_{\mathbf{q}} + \tfrac{1}{2})] \end{aligned} \tag{7.119}$$

So the thermal factor S_T has the form

$$S_T = \sum_{\mathbf{q}}\left(\frac{M(\mathbf{q})}{\omega_{\mathbf{q}}}\right)^2[1 - \cos(\mathbf{q}\cdot\boldsymbol{\delta})](2N_{\mathbf{q}} + 1) \tag{7.120}$$

$$N_{\mathbf{q}} = \frac{1}{e^{\beta\omega_{\mathbf{q}}} - 1} \tag{7.121}$$

The factor S_T is temperature dependent and increases with increasing temperature. For temperatures larger than the Debye temperature, one may often use the expansion $N_{\mathbf{q}} \approx k_B T/\omega_{\mathbf{q}} - \frac{1}{2} + O(\beta\omega_{\mathbf{q}})$ to write

$$S_T \approx 2k_B T\sum_{\mathbf{q}}\left(\frac{M(\mathbf{q})}{\omega_{\mathbf{q}}}\right)^2[1 - \cos(\mathbf{q}\cdot\boldsymbol{\delta})] + O(1/T) \tag{7.122}$$

so that S_T increases linearly with temperature at high temperature. The factor $\exp(-S_T)$ determines the rate of diagonal transitions and decreases with increasing temperature. The Fröhlich-type polaron effects, with coherent motion, are most likely to be found at lower temperatures. As the temperature is increased, the polaron bandwidth $zJ\exp(-S_T)$ becomes smaller and smaller. It is reasonable to expect that the type of motion will change at some temperature. In this model, the change is to a hopping type motion.

The factor δ is the distance the polaron hops in a single motion. It is presumably a lattice constant, or at least a distance associated with some fundamental unit of length on the scale of the crystal unit cell. The factor $[1 - \cos(\mathbf{q} \cdot \boldsymbol{\delta})]$ suggests that phonons of short wavelength are important, since this factor vanishes for $\mathbf{q} \to 0$. The integrations in (7.120) and (7.122) probably have significant contributions from phonons near the edge of the Brillouin zone. These integrals must be done carefully and probably numerically. Simple models, such as the Debye model, are probably very inaccurate. Of course, the matrix elements $M(\mathbf{q})$ must also be found accurately for zone edge phonons. The Fröhlich result $M(\mathbf{q}) \sim 1/q$ is a valid approximation only at long wavelength for LO phonons. It is not adequate for the $\mathbf{q}$ integrations in (7.120) or (7.122).

7.2.4. Nondiagonal Transitions

Nondiagonal transitions are hops, from site $\mathbf{R}_j$ to $\mathbf{R}_j + \boldsymbol{\delta}$, in which the number of phonons is not conserved. It is an inelastic scattering process. The matrix element $\langle f|X_{j+\delta}^{\dagger}X_j|i\rangle$ describes the transition from one quantum state to another. By itself, it does not convey meaningful information. Instead, the transition rate w must be evaluated. It is defined as the rate per unit time at which the hopping occurs. It is calculated by first finding the correlation function $W(t)$, obtained by summing over all final states except the state $|f\rangle = |i\rangle$,

$$W(t) = J^2 \sum_{f \neq i} \langle i|X_j^{\dagger}(t)X_{j+\delta}(t)|f\rangle\langle f|X_{j+\delta}^{\dagger}X_j|i\rangle \tag{7.123}$$

$$= J^2[\langle i|X_j^{\dagger}(t)X_{j+\delta}(t)X_{j+\delta}^{\dagger}X_j|i\rangle - \langle i|X_j^{\dagger}(t)X_{j+\delta}(t)|i\rangle\langle i|X_{j+\delta}^{\dagger}X_j|\rangle] \tag{7.124}$$

and then evaluating the Fourier transform:

$$w = \frac{1}{\hbar^2}\int_{-\infty}^{\infty} dt W(t) \tag{7.125}$$

This formula is just the Fermi golden rule (Schiff, 1955). The transition rate is directly proportional to the dc conductivity calculated from the Kubo formula. The latter relationship is $\sigma = \beta e^2\delta^2 n_0 w/3$. The derivation of σ is provided later, after the evaluation of w. The first term in $W(t)$ is a time-dependent correlation function, which is identical in mathematical form to that found previously for the optical absorption in the many-boson model [see (4.308)]. The previous result may be utilized, by direct analogy, to give the following result:

$$\begin{aligned} \exp[-\Phi(t)] &= \langle i|X_j^{\dagger}(t)X_{j+\delta}(t)X_{j+\delta}^{\dagger}X_j|i\rangle \\ \Phi(t) &= \sum_{\mathbf{q}} \frac{M(\mathbf{q})^2}{\omega_{\mathbf{q}}^2}|1 - e^{i\mathbf{q}\cdot\boldsymbol{\delta}}|^2[(N_{\mathbf{q}} + 1)(1 - e^{-i\omega_{\mathbf{q}}t}) + N_{\mathbf{q}}(1 - e^{i\omega_{\mathbf{q}}t})] \end{aligned} \tag{7.126}$$

The second term in $W(t)$ arises from the diagonal transition rate. Note that $\langle i|X_j^{\dagger}(t)X_{j+\delta}(t)|i\rangle$ is not actually dependent upon time. Repeat the above evaluation of S_T with time-dependent factors λ. Since the result depends upon $|\lambda|^2$, the time cancels out. So the first impulse is to set this term equal to the square of the amplitude, or $J^2\exp(-2S_T)$. This step is too hasty. The thermal average of $|\langle i|X_j^{\dagger}(t)X_{j+\delta}(t)|i\rangle|^2$ is not usually equal to the square of the average. The diagonal transition matrix element, for a state with n phonons, is

$$\langle n|X_{j+\delta}^{\dagger}X_j|n\rangle = e^{-|u_{\mathbf{q}}|^2/2}L_n(|u_{\mathbf{q}}|^2) \tag{7.127}$$

$$u_{\mathbf{q}} = \frac{M(\mathbf{q})}{\omega_{\mathbf{q}}}(1 - e^{i\mathbf{q}\cdot\boldsymbol{\delta}}) \tag{7.128}$$

where $L_n(x)$ is the Laguerre polynominal. The quantity $\exp(-S_T)$ is obtained by averaging this amplitude over all possible values of n. A slightly different result is obtained if we average the square of the amplitude over all possible values of n:

$$|\langle i|X_{j+\delta}^{\dagger}X_j|i\rangle|^2 = \Pi_{\mathbf{q}}e^{-|u_{\mathbf{q}}|^2}[1-e^{-\beta\omega_{\mathbf{q}}}]\sum_{n=0}^{\infty}e^{-\beta\omega_{\mathbf{q}}n}L_n(|u_{\mathbf{q}}|^2)^2 \tag{7.129}$$

$$= \Pi_q I_0\left[2|u_{\mathbf{q}}|^2\sqrt{N_{\mathbf{q}}(N_{\mathbf{q}}+1)}\right]\exp(-2S_T) \tag{7.130}$$

which should be compared to the summation in (7.120). Note that $|\lambda_{\mathbf{q}}| = |u_{\mathbf{q}}|$. The average over the square contains another factor, a modified Bessel function, which is obtained for each state $\mathbf{q}$. The final diagonal hopping probability is obtained by taking the product of this result over all $\mathbf{q}$ states.

Now take the limit where the volume of the solid ν goes to infinity. The particle–phonon matrix element $M(\mathbf{q})$ actually has a factor of $1/\sqrt{\nu}$. Write this matrix element as $M(\mathbf{q}) = \bar{M}(\mathbf{q})/\sqrt{\nu}$, where $\bar{M}$ does not have any volume dependence and is only a function of $\mathbf{q}$. In the limit $\nu \to \infty$, the summations which determine S_T and $\Phi(t)$ become integrals which are well behaved. However, the Bessel function I_0 in (7.130) has an argument which vanishes in this limit. When $z \to 0$, the Bessel function is

$$\lim_{z\to 0} I_0(z) = 1 + \frac{z^2}{4} + O(z^4) \tag{7.131}$$

$$\lim_{z\to 0} \ln[I_0(z)] = \frac{z^2}{4} + O(z^4) \tag{7.132}$$

$$\lim_{\nu\to\infty}\sum_{\mathbf{q}}\ln\left[I_0\left(\frac{\theta(\mathbf{q})}{\nu}\right)\right] = \frac{1}{4\nu}\int\frac{d^3q}{(2\pi)^3}\theta(\mathbf{q})^2 \to 0 \tag{7.133}$$

In the limit of $\nu \to \infty$, the factor of $I_0(z)$ in (7.130) goes to one. The function $W(t)$ is then

$$W(t) = J^2e^{-2S_T}[e^{\phi(t)}-1] \tag{7.134}$$

$$2S_T = \int\frac{d^3q}{(2\pi)^3}\frac{\bar{M}(\mathbf{q})^2}{\omega_{\mathbf{q}}^2}|1-e^{i\mathbf{q}\cdot\boldsymbol{\delta}}|^2(2N_{\mathbf{q}}+1) \tag{7.135}$$

$$\phi(t) = 2S_T - \Phi(t) \tag{7.136}$$

$$= \int\frac{d^3q}{(2\pi)^3}\frac{\bar{M}(\mathbf{q})^2}{\omega_{\mathbf{q}}^2}|1-e^{i\mathbf{q}\cdot\boldsymbol{\delta}}|^2[(N_{\mathbf{q}}+1)e^{-i\omega_{\mathbf{q}}t}+N_{\mathbf{q}}e^{i\omega_{\mathbf{q}}t}] \tag{7.137}$$

The time integral must be done next to obtain the final transition rate in (7.125).

$$w = \frac{J^2}{\hbar^2}e^{-2S_T}\int_{-\infty}^{\infty}dt(e^{\phi(t)}-1) \tag{7.138}$$

which is the Golden Rule of quantum mechanics.

In evaluating this expression, some assumptions have to be made regarding the nature of the spectrum of phonons. One case is an Einstein model where all phonons have the same energy. This case is interesting because there are many combinations of phonons which give the net same change in energy. This case was discussed by Nagaev (1963) but is not treated here. Instead, the phonons will be assumed to possess a realistic spectrum in that most of the phonon states $\mathbf{q}$ have different energies $\hbar\omega_{\mathbf{q}}$. Of course, in a solid the states in the star of the wave vector have the same energy, so there are always several $\mathbf{q}$ states with the same energy.

But the special features from Einstein phonons do not occur from a few states, but from large sections of the Brillouin zone which are degenerate. The restriction to dispersive phonons is asssumed because it is complicated to include events in which energy can be conserved while phonons are being destroyed in one state, say $\mathbf{q}_1$, while being created in another, $\mathbf{q}_2$. Energy is conserved easily if $\omega_{\mathbf{q}_1} = \omega_{\mathbf{q}_2}$ but not otherwise.

For a general spectrum of dispersive phonons, containing both acoustical and optical types, the exponential function $\phi(t)$ has a complicated time dependence, and the time integrals should be done numerically. An approximate result may be obtained by a saddle-point integration, which is done below. First, it is useful to discuss the convergence and properties of the time integral.

At very large values of t, then $\phi(t)$ approaches zero, which is fortunate since otherwise the integral would diverge. However, in one dimension DeWit (1968) has shown that the approach to unity is too slow. He showed that $\lim_{t\to\infty} \phi(t) \to O(1/\sqrt{t})$, so that the integral in (7.138) does not converge. Since there are no one-dimensional solids, perhaps this anomaly is no disaster. However, from the pedagogical viewpoint, it is unfortunate since one-dimensional problems are nice for homework. In three dimensions the convergence is fast enough, so that w has a definite value.

The physics is understood by investigating the frequency spectrum of just the first factor in $W(t)$. Call its Fourier transform $U(\omega)$:

$$U(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} e^{-\Phi(t)} = e^{-2S_T} \int_{-\infty}^{\infty} dt e^{i\omega t + \phi(t)} \tag{7.139}$$

A typical spectrum is shown in Fig. 7.5. There is a delta function at $\omega = 0$ plus a smooth spectrum. The delta function is the *no-phonon transition*. It is the probability that the transition takes place without any phonons being emitted or absorbed. As long as this event has a nonzero probability, one will see a delta function in the spectrum. The intensity of the delta function is just proportional to $\exp(-2S_T)$, which is called the *Debye–Waller factor*. In the hopping transition, the equivalent of the no-phonon line is hopping by a diagonal transition. The diagonal terms should be eliminated. The subtraction of the -1 factor from the time integrand, in the definition of w, is just what is needed to eliminate the delta function. The value of w is just the smooth part of the curve in Fig. 7.5 evaluated at $\omega = 0$.

An approximate result for w may be obtained by a saddle-point integration (Schotte, 1966). This approximate result is very accurate in the limit where the particle–phonon coupling is large, so the polaron effects are significant.

First we rewrite $\phi(t)$ in (7.137) as

$$\phi(t) = 2 \sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \sqrt{N_{\mathbf{q}}(N_{\mathbf{q}} + 1)} \cos\left[\omega_{\mathbf{q}}\left(t + \frac{i\beta}{2}\right)\right] \tag{7.140}$$

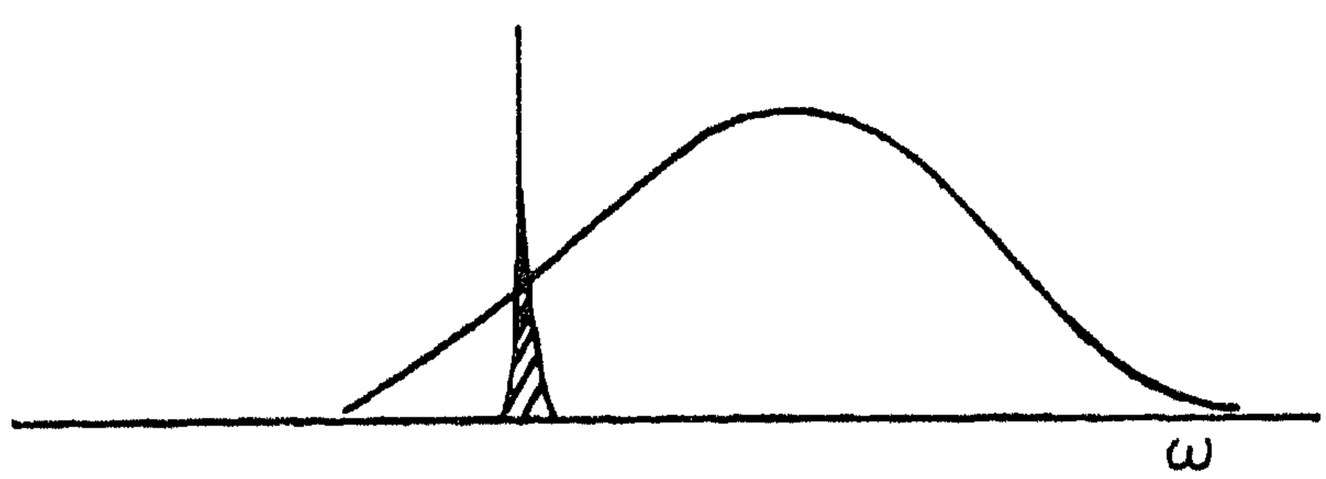


FIGURE 7.5 The function $U(\omega)$.

The t integration is treated as a contour integral in a complex space. The path of integration is deformed to go over the saddle point. The saddle point t_s is located at the point $t_s = -i\beta/2$. The deformed contour is shown in Fig. 7.6, with the saddle point at $\oplus$. The function $\phi(t) = \bar{\phi}(z)$ is expanded as a power series about the saddle point, where the distance from the saddle point is $z = t + i\beta/2$

$$\bar{\phi}(z) = 2\sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \sqrt{N_{\mathbf{q}}(N_{\mathbf{q}}+1)}\cos(\omega_{\mathbf{q}} z) \tag{7.141}$$

$$\approx \bar{\phi}(0) - \gamma^2 z^2 + O(z^4) \tag{7.142}$$

$$\bar{\phi}(0) = 2\sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \sqrt{N_{\mathbf{q}}(N_{\mathbf{q}}+1)} \tag{7.143}$$

$$\gamma = \sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \omega_{\mathbf{q}}^2 \sqrt{N_{\mathbf{q}}(N_{\mathbf{q}}+1)} \tag{7.144}$$

An approximate result is obtained by neglecting all terms except the first two. The time integration, in the vicinity of the saddle point, is now just a Gaussian integral:

$$\int dz e^{\bar{\phi}(z)} \approx e^{\bar{\phi}(0)} \int dz e^{-\gamma z^2} = e^{\bar{\phi}(0)} \sqrt{\frac{\pi}{\gamma}}$$

$$w = \frac{J^2}{\hbar^2}\sqrt{\frac{\pi}{\gamma}} \exp[-2S_T + \bar{\phi}(0)] \tag{7.145}$$

The saddle-point integral does the $\exp[-\phi(t)]$ part of the integrand and not the -1 term. However, the latter contribution just eliminates the delta function corresponding to the no-phonon transitions. The saddle-point integration does not have this contribution. It provides the smooth background part of $U(\omega)$ at $\omega = 0$, which is precisely the contribution which is needed for w.

The two exponential terms in (7.145) can be combined,

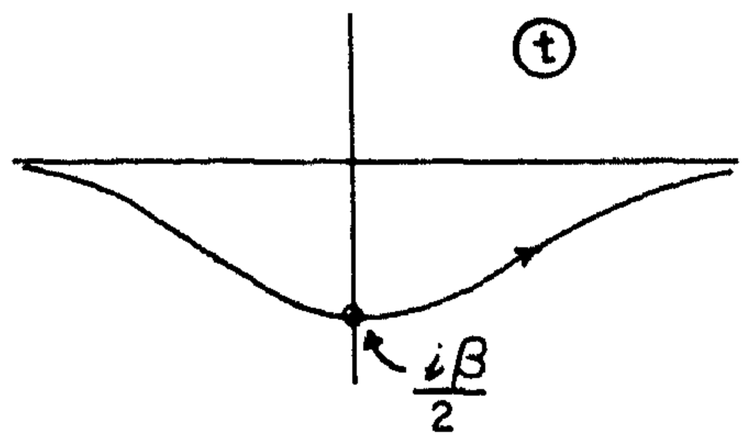
$$2S_T - \bar{\phi}(0) = \sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \left[\sqrt{N_{\mathbf{q}}+1} - \sqrt{N_{\mathbf{q}}}\,\right]^2 \tag{7.146}$$

which is a positive definite form. The exponent in (7.145) is always negative. At high temperatures, the phonon occupation numbers may be expanded: $N_{\mathbf{q}} \to (k_B T/\omega_{\mathbf{q}}) - \frac{1}{2} + O(\omega_{\mathbf{q}}/k_B T)$. Define $x = \beta\omega_{\mathbf{q}}$, and the factors in the exponent give

$$\lim_{x\to 0}\left[\sqrt{(N_{\mathbf{q}}+1} - \sqrt{N_{\mathbf{q}}}\right]^2 = 2N_{\mathbf{q}} + 1 - 2\sqrt{N_{\mathbf{q}}(N_{\mathbf{q}}+1)}$$

$$= \frac{2}{x} - 2\left(\frac{1}{x^2} - \frac{1}{4}\right)^{1/2} = \frac{x}{4} + O(x^3) \tag{7.147}$$

$$2S_T - \bar{\phi}(0) \approx \frac{1}{4k_B T}\sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \omega_{\mathbf{q}} \tag{7.148}$$



The high-temperature expansion defines an activation energy $\bar{\Delta}$ for the hopping rate

$$\bar{\Delta} = \tfrac{1}{4}\sum_{\mathbf{q}} |u_{\mathbf{q}}|^2 \omega_{\mathbf{q}} = \tfrac{1}{4}\int \frac{d^3q}{(2\pi)^3} \frac{\bar{M}(\mathbf{q})^2}{\omega_{\mathbf{q}}} |1 - e^{i\mathbf{q}\cdot\boldsymbol{\delta}}|^2 \tag{7.149}$$

The same factor $\bar{\Delta}$ defines $\gamma = 4\bar{\Delta}k_B T$ at high temperature.

Small-polaron theory predicts that the hopping rate is thermally activated. The rate of hopping increases at higher temperatures. This behavior is in direct contrast to the rate $J^2 \exp(-2S_T)$ of diagonal transitions, which decreases with increasing temperatures. As temperature increases, the diagonal transitions become less likely, while the nondiagonal transitions become more likely. The band-type polaron motion, or large polarons of the Fröhlich type, will exist only at small temperatures. Holstein estimated the transition temperature between band motion and hopping motion to occur around 40% of the Debye temperature. The estimate is remarkably insensitive to the magnitude of J. In the Holstein model, the low-temperature motion should be band-like, while the high-temperature motion is hopping. There have been many experimental systems with these characteristics which have been ascribed to small-polaron theory. One example is TiO_2 (Bogomolov *et al.*, 1968). They observe the transition from band to hopping conductivity at about 300°C. The conduction bands of transition metal oxides are often d bands, and conduction in them seems well described by small-polaron theory. See the reviews by Adler (1967), Appel (1967), and Böttger and Bryksin (1976).

There are many other systems in which the electron conductivity is thermally activated. They may often be otherwise explained, e.g., by an activation energy for freeing bound electrons from defects or alternately by a static barrier to the electron motion. However, it is thought that in most systems in which the electron motion is by hopping there is some polarization of the electron by its immediate surroundings. The hopping electron must carry its local polarization along, which gives the activation energy. The small-polaron picture applies when this polarization is due to phonons or atomic realignment near the electron. Of course, there could also be polarization of the electronic states, for example, the dielectric screening by the material of the electron charge. Electronic polarization can be described by an *electronic polaron model*, which has the same mathematical form as small-polaron theory. Here the bosons are not phonons but density operators representing electron–hole pair excitations of the system. This concept applies even in insulators, where the pair oscillations have an energy gap.

The important message in small-polaron theory is that the motion is thermally activated. If the particle polarizes its surroundings, then it can hop only by moving this polarization along. The greater the polarization, the less likely hopping occurs. Each jump may occur only when the polarization arrangements on initial and final sites are the same. Since the system is in a state of continual fluctuation, this coincidence sometimes does happen. It is less likely when the polarization is most severe. That is the reason for the thermal activation.

7.2.5. Kubo Formula

The Kubo formula for small polarons was evaluated by Lang and Firsov (1963, 1964). The conductivity $\sigma(\omega)$ is evaluated from the current–current correlation function. The limit of $\omega \to 0$ provides the dc conductivity, while the optical absorption is given by the result for

nonzero frequencies. Both results are interesting for small polarons. The discussion will begin with the form of the Kubo formula given in Problem 16 of Chapter 3:

$$\mathrm{Re}(\sigma) = \frac{1 - e^{-\beta\omega}}{2\omega} \int_{-\infty}^{\infty} dt e^{i\omega t} \langle \mathscr{J}_\alpha^\dagger(t) \mathscr{J}_\alpha(0) \rangle \tag{7.150}$$

$$\mathscr{J}_\alpha = i\frac{Je}{\hbar} \sum_{j\delta} \delta_\alpha C_{j+\delta}^\dagger C_j X_{j+\delta}^\dagger X_j \tag{7.151}$$

The electrical current operator is the hopping form (1.204), which has been subsequently altered by the canonical transformation (7.108). The spatial subscripts α are replaced by the diagonal sum in cubic crystals:

$$\mathrm{Re}[\sigma(\omega)] = \frac{1 - e^{-\beta\omega}}{6\omega} \frac{J^2 e^2}{\hbar^2} \sum_{jj'\delta\delta'} \boldsymbol{\delta} \cdot \boldsymbol{\delta}' \int_{-\infty}^{\infty} dt e^{i\omega t}$$
$$\times \langle C_{j+\delta}^\dagger(t) C_j(t) C_{j'+\delta'}^\dagger C_{j'} \rangle \langle X_{j+\delta}^\dagger(t) X_j(t) X_{j'+\delta'}^\dagger X_{j'} \rangle \tag{7.152}$$

In the electron correlation function, the operators have a time development governed by $e^{i\bar{H}t} C e^{-i\bar{H}t}$, where $\bar{H}$ also contains the hopping term V in (7.112). In the interaction representation, the expansion of the S-matrix will result in terms of higher power in J. These terms are seldom considered, and it is customary to evaluate only the term of order J^2, which is certainly the leading term. However, the neglect of higher terms in J is mostly expediency, since they are hard to calculate. There is generally no proof available that these higher-order terms are smaller.

In the interaction representation, the zeroth-order term in the S-matrix expansion just replaces the electron part of the correlation function by its value obtained using the time development governed by H_0 rather than $\bar{H}$. This leading term in the conductivity will be called $\sigma^{(0)}(\omega)$. In this approximation, the electron correlation function is easy to evaluate since there is no time dependence:

$$\langle C_{j+\delta}^\dagger(t) C_j(t) C_{j'+\delta'}^\dagger C_{j'} \rangle = \delta_{j=j+\delta'} \delta_{\delta=-\delta'} \langle n_{j'}(1 - n_{j'+\delta'}) \rangle \tag{7.153}$$

The correlation function is just equal to $c(1 - c)$, which is the probability c that the initial site is occupied times the probability $(1 - c)$ that the final site is empty. It is the correct average for this correlation function whenever there is no correlation between the site occupations of neighboring particles. Such correlations exist, for example, whenever the forces between particles on different sites are included. Most model calculations omit such forces.

The electron correlation function is nonzero only when $j = j' + \delta'$ and $\boldsymbol{\delta} = -\boldsymbol{\delta}'$. The calculation of the real part of the conductivity $\sigma^{(0)}(\omega)$ is now reduced to the evaluation of the correlation function for phonon coordinates, which is the nondiagonal transition rate for each site, summed over sites

$$\mathrm{Re}[\sigma^{(0)}(\omega)] = \frac{1 - e^{-\beta\omega}}{6\omega} \frac{J^2 e^2}{\hbar^2} c(1 - c) \sum_{j\delta} \delta^2 \int_{-\infty}^{\infty} dt e^{i\omega t} \langle X_j^\dagger(t) X_{j+\delta}(t) X_{j+\delta}^\dagger X_j \rangle \tag{7.154}$$

The solid is assumed to have sufficient symmetry that the rate for each pair of sites is the same. The summation over sites j just yields the number of sites, which is called N. The summation over δ yields the coordination number z. The number of particles is $N_e = Nc$,

where c is the concentration. For dimensional reasons, divide by the volume of the solid ν, so that the real part of the conductivity is

$$\mathrm{Re}[\sigma^{(0)}(\omega)] = \frac{1-e^{-\beta\omega}}{6\omega}\left(\frac{Nz}{\nu}\right)\frac{J^2e^2\delta^2}{\hbar^2}c(1-c)U(\omega) \tag{7.155}$$

$$U(\omega) = \int_{-\infty}^{\infty} dte^{i\omega t}\langle X_j^\dagger(t)X_{j+\delta}(t)X_{j+\delta}^\dagger X_j\rangle \tag{7.156}$$

The conductivity is proportional to the function $U(\omega)$, which was mentioned above. It has a delta function at zero frequency, which should also be eliminated from $\mathrm{Re}[\sigma^{(0)}(\omega)]$. The real part of the conductivity $\sigma^{(0)}(\omega)$ is just proportional to the Fourier transform of the hopping correlation function $W(t)$. The dc result $\omega \to 0$ was evaluated earlier in (7.95) by a saddle-point integration, which for the dc conductivity gives

$$\sigma_{dc}^{(0)} = c(1-c)\left(\frac{zNe^2}{6\nu}\right)\beta w \tag{7.157}$$

$$w = \frac{J^2}{\hbar^2}\left(\frac{\pi}{\gamma}\right)^{1/2}\exp\left[-\frac{\bar{\Delta}}{k_BT}\right] \tag{7.158}$$

The same saddle-point integration also provides an estimate of the frequency-dependent conductivity (Reik, 1972). Again the integration variable is changed to $z = t + i\beta/2$, and $\bar{\phi}(z)$ is expanded about the point $z = 0$:

$$U(\omega) = e^{-2S_T}\int_{-\infty}^{\infty} dte^{i\omega t}e^{\phi(t)} \tag{7.159}$$

$$= e^{-2S_T+\beta\omega/2}\int_{-\infty}^{\infty} dze^{i\omega z+\bar{\phi}(z)} \tag{7.160}$$

$$\approx e^{-2S_T+\beta\omega/2+\bar{\phi}(0)}\int_{-\infty}^{\infty} dze^{i\omega z-\gamma z^2} \tag{7.161}$$

$$\approx \left(\frac{\pi}{\gamma}\right)^{1/2}\exp\left[-2S_T+\beta\omega/2+\bar{\phi}(0)-\frac{\omega^2}{4\gamma}\right] \tag{7.162}$$

$$\mathrm{Re}[\sigma^{(0)}(\omega)] = \sigma_{dc}^{(0)}\frac{\sinh(\beta\omega/2)}{(\beta\omega/2)}\exp\left[-\frac{\omega^2}{4\gamma}\right] \tag{7.163}$$

At nonzero frequency the integral over z is still of the Gaussian form and yields a Gaussian function of frequency. The result for $\mathrm{Re}[\sigma^{(0)}(\omega)]$ is important for relating theory and experiment. It provides a prediction of proportionality between the dc conductivity and the optical absorption. For example, it predicts that the optical absorption is Gaussian, with a width given by $4\gamma = 16\bar{\Delta}k_BT$, where $\bar{\Delta}$ is the activation energy observed in the dc conductivity. This prediction may be tested experimentally. It also predicts that the magnitude of the absorption is proportional to the magnitude of the dc conductivity at each temperature. These predictions are confirmed in TiO_2. Figure 7.7 shows the optical data of Kudinov *et al.* (1970). The absorption data is in satisfactory agreement with the theoretical curve, which is calculated using parameters which also fit the dc conductivity. The physics of the absorption process may be understood using a configurational coordinate picture suggested by Polder (unpublished; see Reik, 1972). The coordinates are illustrated in Fig. 7.8. A particle sitting at site j has a parabolic potential energy curve which represents the potential energy of the

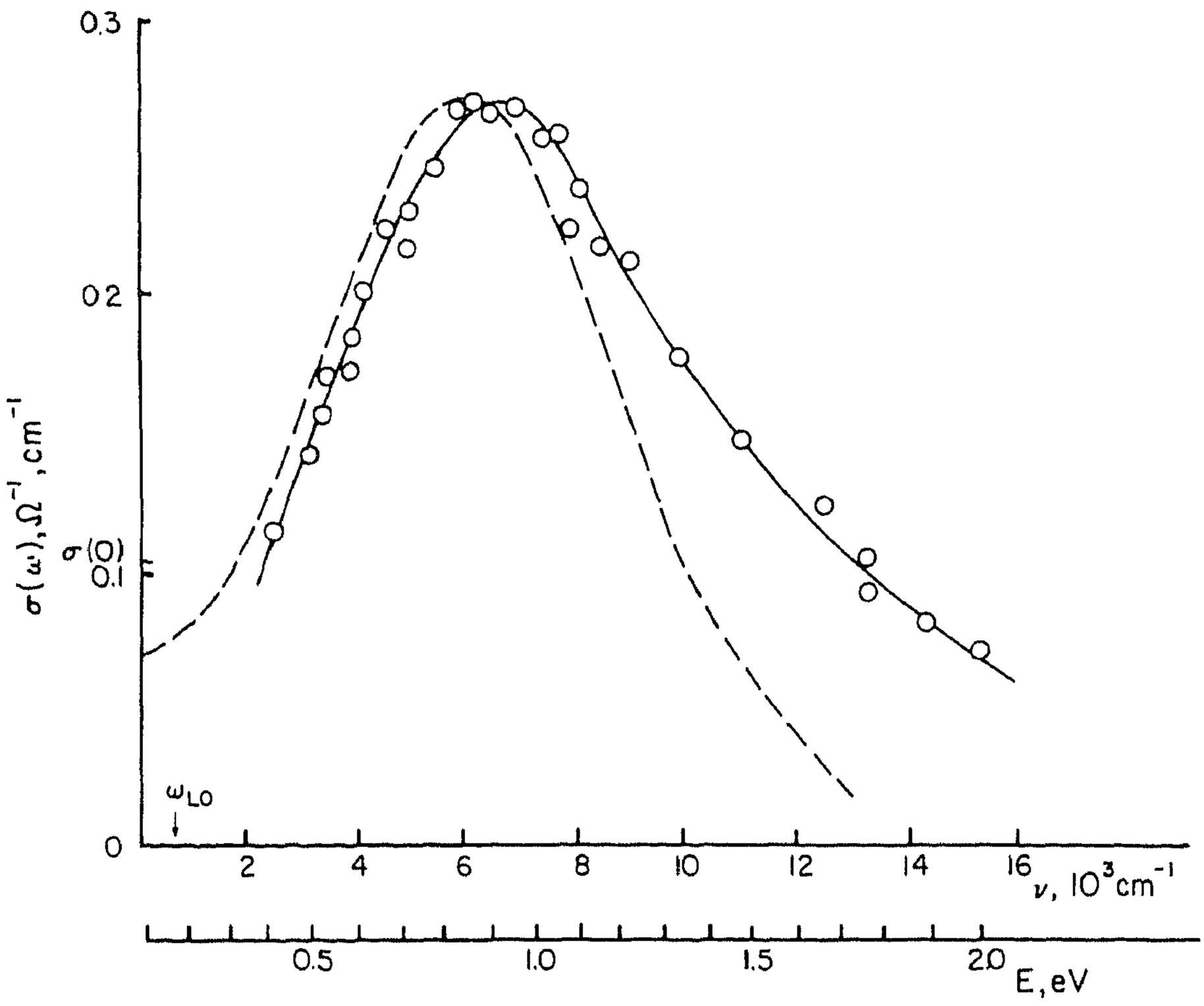


FIGURE 7.7 Optical absorption by small polarons in TiO_2 for $E \| c$ at $T = 600$ K given by points. Dashed curve is theory with $\hbar\omega_0 = 800\ \text{cm}^{-1} = 0.1$ eV. *Source*: Kudinov *et al.* (1970) (used with permission).

phonons for a displacement Q about the equilibrium point. The neighboring site $j + \delta$ has an identical parabola, displaced to the new equilibrium point, at a distance Q_0. The dc conductivity occurs by thermal activation over the intervening potential barrier $\bar{\Delta}$, which is midway between the two parabolas. If the parabolas have a curvature given by α, so that the potential energy curves are $V(Q) = \alpha Q^2$, then the activation energy is $\bar{\Delta} = V(Q_0/2) = \alpha Q_0^2/4$. Similarly, the most probable optical transition is the vertical arrow in Fig. 7.8. It puts the particle on the neighboring parabola, and later the particle relaxes down to the minimum at the neighboring point. The length of this arrow is the energy $\hbar\omega_m = V(Q_0) = \alpha Q_0^2$, which is four times the activation energy. The proportionality $\hbar\omega_m = 4\bar{\Delta}$ agrees with the result (7.163). The configurational coordinate model may be used, in fact, to derive (7.163) directly.

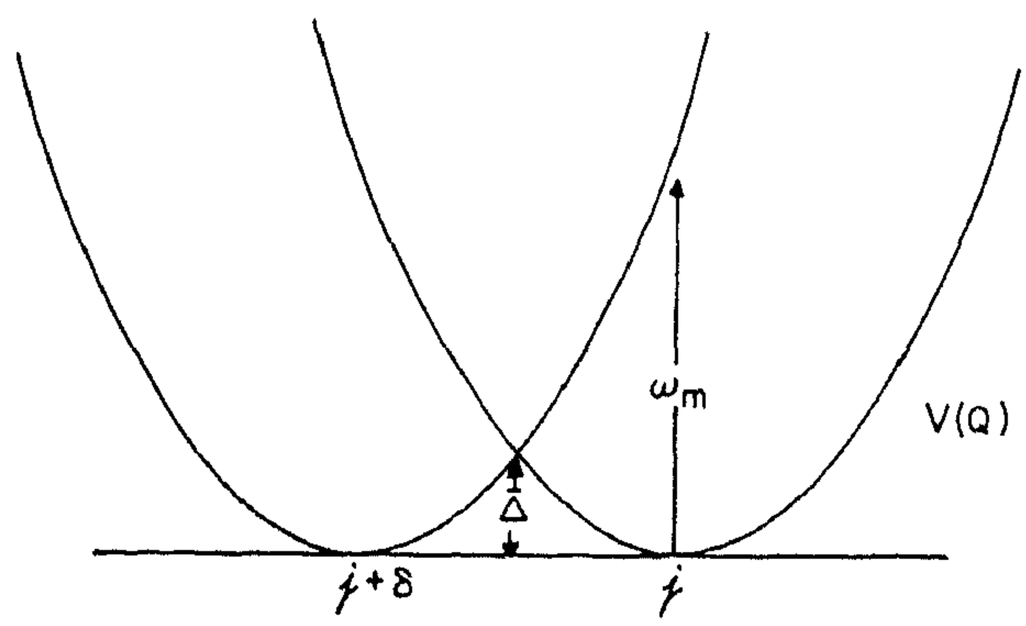


FIGURE 7.8 Configurational coordinate drawing of small-polaron motion. The probable optical absorption is the transition ω_m, while the hopping barrier is $\bar{\Delta}$.

7.3. HEAVILY DOPED SEMICONDUCTORS

An interesting application of polaron theory is in many-particle systems. If the solid has many electrons, one has to consider electron–electron interactions along with the electron–phonon interactions. They are very important in metals, which are discussed in the next section. Other systems, which mix the two theories of polarons and electron–electron interactions, are semiconductors which are heavily doped with impurities. Many III–V semiconductors are polar and may also be doped sufficiently with impurities to form a degenerate electron gas. Then at low temperatures the electron gas has a well-defined Fermi energy, and the electrical resistivity is finite.

These systems are interesting from several viewpoints. As an electron gas, they may have an effective density which is very high. The actual density is quite low, perhaps only $n_0 \sim 10^{18}-10^{19}$ cm^{-3}. However, the value of r_s is calculated with the effective Bohr radius,

$$a_0^* = \frac{\hbar^2 \varepsilon_0}{e^2 m^*} \tag{7.164}$$

$$r_s = \left(\frac{4\pi n_0 a_0^{*3}}{3}\right)^{-1/3} \tag{7.165}$$

which has the effective mass m^* and static dielectric function ε_0 for the semiconductor. Because typical values are $m^*/m \sim 0.1-0.2$ and $\varepsilon_0 \sim 10$, the effective Bohr radius may be as large as 5 nm. Even for an electron concentration of $n_0 \sim 10^{18}$ cm^{-3}, a Bohr radius of $a_0^* \sim 5$ nm gives $r_s \sim 1$, which is a smaller value than found in metals. Even smaller values of r_s are obtainable in many semiconductors. Semiconductors are a suitable environment in which to study the high-density electron gas.

Many semiconductors are polar. The interaction between electrons in the conduction band and LO phonons is well described by the Fröhlich Hamiltonian. Many of these semiconductors may also be doped to high electron concentration, so that experimentally one may study polaron theory in a many-particle system. However, nature does not seem to like these two conditions—strong polaron interactions and high doping levels—to occur in the same materials. Materials which are very polar, such as alkali halides or II–VI semiconductors, can seldom be doped to where the electron gas is degenerate. Group IV (Si, Ge) and III–V semiconductors may be doped readily to high levels but are either weakly polar or not polar. The theory of polarons, in many-particle semiconductors, needs only to treat systems with weak coupling between electrons and phonons. An advantage of these experimental systems is that all the parameters of the Hamiltonian are known from other measurements. The only parameters of the theory are m^*, n_0, ε_0, and ε_∞, which all can be obtained by simple measurements. Many-body theory may be tested in situations where there are no adjustable parameters.

The self-energy of an electron from electron–phonon interactions is dramatically altered by the presence of other electrons. Several effects are described. First, the phonon energies themselves are altered by the screening properties of the electron gas. Second, the electron–phonon interaction is screened by the electron gas. The screening is a significant feature, so that even intermediate values of α are reduced to weak coupling. Third, the nature of the electron's self-energy is changed dramatically by the Pauli exclusion principle. When electrons interact with phonons, they can only scatter into states not occupied by other electrons. The sum of these three effects is that polaron theory is entirely changed from the simple one-particle theory of the previous sections.

7.3.1. Screened Interaction

A little thought shows the system is complicated. The phonons and the electron gas are mutually interacting and affecting each other's properties. Should the first step, in the calculation, be to calculate the effect of the phonons on the electrons, or vice versa? The answer is neither. Instead, one should calculate the total longitudinal dielectric function of the system, which includes parts from both electrons and phonons. This function will be sufficient to describe all the relevant physics.

The effective interaction between electrons may be written as

$$V_{\text{eff}}(q, \omega) = \frac{4\pi e^2}{q^2 \varepsilon_{\text{total}}(q, \omega)} = \frac{v_q}{\varepsilon_{\text{total}}(q, \omega)} \tag{7.166}$$

where $\varepsilon_{\text{total}}(q, \omega)$ is the total dielectric function from all sources. The total dielectric function will be derived in three ways. The first will be a phenomenological derivation. The other two derivations use Green's functions.

The simple derivation is given first. The dielectric function is the summation of the three contributions: ε_∞, the electron–electron interactions, and the electron–phonon interactions. The first two are known, so that

$$\varepsilon_{\text{total}}(q, \omega) = \varepsilon_\infty - v_q P(q, \omega) + \text{phonon term} \tag{7.167}$$

where $P(q, \omega)$ is the polarization of the electron gas, which was discussed in Sec. 5.5. The phonon term will be calculated assuming that there are no electrons present.

For optical phonons in polar crystals, the ions will vibrate in response to an oscillating electric field. If X is the distance between a pair of plus and minus ions with reduced mass M, the classical equations of motion are

$$M\ddot{X} + KX = eE_0 e^{-i\omega t} = eE(t) \tag{7.168}$$

The force constant is $K = M\omega_{TO}^2$. The oscillatory solution has a periodic displacement of

$$X(t) = \frac{eE(t)}{M} \frac{1}{\omega_{TO}^2 - \omega^2} \tag{7.169}$$

The net polarization of the system is $P = en_0 X$, where e and n_0 are the charge and density of the ion pairs. The contribution to the dielectric function is

$$\Delta\varepsilon = 4\pi\alpha = 4\pi e n_0 \frac{X(t)}{E(t)} = \frac{4\pi e^2 n_0}{M} \frac{1}{\omega_{TO}^2 - \omega^2} \tag{7.170}$$

The collection of constants $4\pi e^2 n_0/M$ is an ion plasma frequency. It may be determined phenomenologically by realizing that in the absence of the electron–electron term, the long-wavelength dielectric function of a polar crystal may be written as

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{1 - \omega^2/\omega_{TO}^2} \tag{7.171}$$

$$\varepsilon_0 - \varepsilon_\infty = \frac{4\pi e^2 n_0}{M\omega_{TO}^2} \tag{7.172}$$

At zero frequency $\varepsilon(0) = \varepsilon_0$, while at frequencies above the reststrahl $\omega \gg \omega_{TO}$ then $\varepsilon(\omega) = \varepsilon_\infty$. These limits define the two quantities ε_0 and ε_∞. The total dielectric function is

$$\varepsilon_{\text{total}}(q, \omega) = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{1 - \omega^2/\omega_{TO}^2} - v_q P(q, \omega) \tag{7.173}$$

In the derivation, the three contributions to the dielectric function are treated as independent and additive. This assumption is not quite right. The rigorous derivation shows that the phonons affect the electron–electron contribution. When calculating the polarization diagrams for $P(q, \omega)$, there are diagrams in which the basic electron bubble has internal phonon lines. However, these terms are never included in actual calculations. Instead, the term $v_q P(q, \omega)$ is approximated by either the RPA or the Thomas–Fermi model. In this approximation, the three terms are independent.

The rigorous derivation of the total screening function proceeds by summing diagrams. The Hamiltonian has the form

$$H = H_0 + H_{ee} + H_{ep} + H_{pp} \tag{7.174}$$

$$H_0 = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \sum_{\mathbf{q}\lambda} \omega_{\mathbf{q}\lambda} a^\dagger_{\mathbf{q}\lambda} a_{\mathbf{q}\lambda} \tag{7.175}$$

$$H_{ee} = \frac{1}{2\nu} \sum_{\mathbf{q}} v_q^{(\infty)} \rho(\mathbf{q})\rho(-\mathbf{q}) \tag{7.176}$$

$$H_{ep} = \frac{1}{\sqrt{\nu}} \sum_{\mathbf{q}\lambda} M_{\mathbf{q}\lambda} \rho(\mathbf{q}) A^\dagger_{\mathbf{q}\lambda} \tag{7.177}$$

$$H_{pp} = \tfrac{1}{2} \sum_{\mathbf{q}\lambda} V_\lambda(\mathbf{q}) A^\dagger_\lambda(\mathbf{q}) A_\lambda(\mathbf{q}) \tag{7.178}$$

$$v_q^{(\infty)} = \frac{v_q}{\varepsilon_\infty}, \qquad A_\lambda(\mathbf{q}) = a_{\mathbf{q}\lambda} + a^\dagger_{-\mathbf{q}\lambda}, \qquad \rho(\mathbf{q}) = \sum_{\mathbf{p}\sigma} C^\dagger_{\mathbf{p}+\mathbf{q}\sigma} C_{\mathbf{p}\sigma} \tag{7.179}$$

The electron density operator is $\rho(\mathbf{q})$. The electron–phonon matrix element is $M_{\mathbf{q}\lambda}$. The phonon–phonon matrix element is $V_\lambda(\mathbf{q})$. The electron–electron interaction has been written as the product of two density operators. This form has the liability that it permits an electron to interact with itself. These terms are meant to be absent and are ignored. The shorthand notation is convenient for the following discussion.

The term H_0 describes the noninteracting systems of electrons and phonons. The electron's energy $\xi_{\mathbf{p}} = \varepsilon_{\mathbf{p}} - \mu$ is measured from its chemical potential. The kinetic energy is usually assumed to have the form $\varepsilon_{\mathbf{p}} = p^2/2m^*$. Band structure effects are included by an effective mass m^*. In semiconductors m^* is usually smaller than a bare electron mass. The electron–phonon part of the Hamiltonian has the usual form. Our applications will mostly be to polar coupling, where $M_{\mathbf{q}\lambda}$ has the form in (7.2). The electron–electron interaction has an interaction $v_q^{(\infty)} = v_q/\varepsilon_\infty$.

The factor ε_∞ is included to account for the dielectric screening of the material. There are three sources of dielectric screening in the system. The first is the electron–electron interactions from the mobile electrons in the conduction band; their contribution to the screening is included explicitly by the term H_{ee}. The second is from the optical phonons, and they are included through the term H_{ep}. The third is from the high-energy electronic excitations across the band gap of the semiconductor. They give rise to ε_∞. They could also be included as an additional term in the Hamiltonian, which would be solved to give ε_∞. However, for the low-frequency excitations of interest in the present discussion, these high-

energy excitations are uninteresting because they give a constant contribution to $\varepsilon(\mathbf{q}, \omega)$. It is simpler to include these effects through the constant ε_∞, which is introduced in the Hamiltonian at the beginning.

There are several different ways to solve these equations which give the same answer. The easiest method is done first. In order to keep the discussion simple, the treatment is specialized to longitudinal optical (LO) phonons. The ions are treated as point masses with an effective charge eZ. The various interaction terms are approximated as

$$\omega_{\mathbf{q}\lambda} = \omega_{TO}, \qquad M_{\mathbf{q}\lambda} = v_q^{(\infty)} ZqX(q) \tag{7.180}$$

$$V_\lambda(\mathbf{q}) = v_q^{(\infty)} Z^2 q^2 X(q)^2, \qquad X(q)^2 = \frac{n_0 \hbar}{2M\omega_{TO}} \tag{7.181}$$

The phonon frequency $\omega_{\mathbf{q}\lambda}$ is chosen to be the TO frequency, rather than the LO frequency. The difference between these frequencies arises from the long-range Coulomb interactions in H_{pp}. As long as the phonon–phonon interactions are explicitly included in the interaction term of the Hamiltonian, then the TO frequency must be used in H_0. Later it is shown how the term H_{pp} changes ω_{TO} to ω_{LO}.

All of the interactions are due to long-range Coulomb interactions: electron–electron, electron–phonon, and phonon–phonon. All of the interaction terms can be combined into one term

$$H_{\text{int}} = H_{ee} + H_{ep} + H_{pp} \tag{7.182}$$

$$= \tfrac{1}{2} \sum_{\mathbf{q}} v_q^{(\infty)} \rho_T^\dagger(\mathbf{q}) \rho_T(\mathbf{q}) \tag{7.183}$$

$$\rho_T(\mathbf{q}) = \frac{\rho(\mathbf{q})}{\sqrt{\nu}} + ZqX(\mathbf{q})A(\mathbf{q}) \tag{7.184}$$

The total density operator $\rho_T(\mathbf{q})$ is the summation of the density operators for electron and phonons.

The dielectric function is calculated for this interaction by following the same steps used in Sec. 5.5.2 to obtain RPA. The difference between that case and the present one is the phonons. By following the same steps, one can show that

$$\varepsilon(\mathbf{q}, i\omega) = \varepsilon_\infty [1 - v_q^{(\infty)} P_T^{(1)}(\mathbf{q}, i\omega)] \tag{7.185}$$

$$P_T^{(1)}(\mathbf{q}, i\omega) = -\int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau \rho_T(\mathbf{q}, \tau) \rho_T^\dagger(\mathbf{q}, 0) \rangle \tag{7.186}$$

The electron and phonon operators act independently. The most important terms in $P^{(1)}$ are the summation of the separate susceptibilities for electrons and phonons:

$$P_T^{(1)}(\mathbf{q}, i\omega) = P_e^{(1)}(\mathbf{q}, i\omega) + P_{ph}^{(1)}(\mathbf{q}, i\omega) \tag{7.187}$$

$$P_{ph}^{(1)}(\mathbf{q}, i\omega) = Z^2 q^2 X(q)^2 \mathscr{D}(\mathbf{q}, i\omega) \tag{7.188}$$

$$-v_q P_{ph}^{(1)}(\mathbf{q}, i\omega) = \frac{\omega_{pi}^2}{\omega_{TO}^2 - \omega^2}, \quad \omega_{pi}^2 = \frac{4\pi e^2 z^2 n_0}{M} \tag{7.189}$$

The electron polarization $P_e^{(1)}$ is identical to the RPA result in Sec. 5.5.2. This dielectric function is identical to the one found in (7.173). It is the RPA result for a system of electrons and LO phonons. This derivation is rather simple because all long-range interactions were included in H_{int}.

The third derivation of this same result is the one presented in most reviews. The first step is derive an effective interaction $W(q)$ between electrons. It has two terms. The first is the direct Coulomb interaction $v_q^{(\infty)}$. The second term is the interaction between two electrons which results by the exchange of a phonon.

$$W(q) = v_q^{(\infty)} + V_{ph}(q) \tag{7.190}$$

$$V_{ph}(q) = M_q^2 D(q) \tag{7.191}$$

The phonon Green's function $\mathscr{D}(q)$ should include the phonon self-energy due to phonon–phonon interaction H_{pp}. This self-energy is $V_\lambda = \omega_{pi}^2/2\omega_{TO}$. The phonon Green's function is

$$D(q, \omega) = \frac{-2\omega_{TO}}{\omega^2 - \omega_{TO}^2 - 2\omega_{TO}V_\lambda} = \frac{-2\omega_{TO}}{\omega^2 - \omega_{LO}^2} \tag{7.192}$$

$$\omega_{LO}^2 = \omega_{TO}^2 + \omega_{pi}^2 \tag{7.193}$$

The RPA expansion for the dielectric function is now evaluated with the effective interaction $W(q)$ which gives that

$$V_{\text{eff}}(q) = \frac{W(q)}{1 - W(q)P^{(1)}(q)} = \frac{v_q}{\varepsilon_{\text{total}}(q)} \tag{7.194}$$

The last equal sign states that this expression is identical to the previous one in terms of the total dielectric function. Proving this identity is assigned as a homework problem.

The total dielectric function is often well approximated by the sum of the independent contributions. For example, the piezoelectric contribution from acoustical phonons provides another term in the above series (Mahan, 1972) for crystals which are piezoelectric. The starting point of a many-body calculation for any system with several contributions should be to find the dielectric function. Often a satisfactory approximation is obtained by just adding the separate contributions. It will be shown later that this does provide a description of the mixing and interference between these modes. The mixing occurs because all these various modes—LO phonons, plasmons, piezoelectric phonons—have longitudinal electric fields. The longitudinal electric fields of the various modes cause the mutual coupling. This coupling is well described by the dielectric function.

A curious feature of the optical phonon part of the dielectric function $(\varepsilon_0 - \varepsilon_\infty)/(1 - \omega^2/\omega_{TO}^2)$ is that the resonance frequency is ω_{TO}. The resonance is the frequency where the energy denominator vanishes. This position of resonance is in contrast to the Green's function for optical phonons, which has its pole or resonance frequency at the LO phonon frequency ω_{LO}. There is no disagreement between these two results. The longitudinal excitations of the system are not at the poles of $\varepsilon(\omega)$ but are at the poles of $1/\varepsilon(\omega)$. For example, if the polar solid had no conduction electrons and was an insulator, the total dielectric function would just be the first two terms of (7.173) or

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{1 - \omega^2/\omega_{TO}^2} \tag{7.195}$$

$$\frac{1}{\varepsilon(\omega)} = \frac{1 - \omega^2/\omega_{TO}^2}{\varepsilon_0 - \varepsilon_\infty \omega^2/\omega_{TO}^2} \tag{7.196}$$

$$= \frac{1}{\varepsilon_\infty} - \frac{\omega_{LO}^2}{\omega_{LO}^2 - \omega^2}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \tag{7.197}$$

$$\omega_{LO}^2 = \omega_{TO}^2 \frac{\varepsilon_0}{\varepsilon_\infty} \tag{7.198}$$

The poles of $1/\varepsilon(\omega)$ are indeed at ω_{LO}. Of course, in a polar solid there are both LO and TO phonons. However, only the LO phonons make long-range electric fields, so only they respond to a longitudinal electric field. The last relationship between the phonon frequencies and dielectric constants is due to Lyddane *et al.* (1941).

The combined system of electrons and phonons has the dielectric function $\varepsilon_{\text{total}}(q, \omega)$ given in (7.173). The excitations of this system must be given by the poles in $1/\varepsilon_{\text{total}}(q, \omega)$. Let $Z_i(q)$ denote the complex values of ω at which these poles are found. According to the theorem of residues, the complex function may be expressed exactly as the summation over these poles and their residues R_i:

$$\frac{1}{\varepsilon_{\text{total}}(q, Z)} = 1 + \sum_i \frac{R_i(q)}{Z^2 - Z_i(q)^2} \tag{7.199}$$

The poles always occur in the pairs $\pm Z_i$ since $\varepsilon_{\text{total}}$ is an even function of ω. Each pole is then interpreted as an excitation of the system, with a coupling strength given by the residue. The expansion in poles and residues would be an exact way to describe the excitations, and their couplings, if $1/\varepsilon_{\text{total}}$ were only a sum of poles. However, previous experience indicates that there are branch cuts in the complex function. In the electron gas, these described the pair excitations, and the same excitations should occur here. The exact description of the excitation spectrum is all contributions where there is a nonzero contribution from

$$S(q, \omega) = \frac{1}{n_0 v_q} \operatorname{Im}\left[\frac{1}{\varepsilon_{\text{total}}(q, \omega + i\delta)}\right] \tag{7.200}$$

which would include poles as well as branch cuts.

There are two important frequencies in the coupled system: the LO phonon frequency ω_{LO} and the plasma frequency of the electron gas:

$$\omega_p^2 = \frac{4\pi e^2 n_0}{\varepsilon_\infty m^*} \tag{7.201}$$

The coupled modes have a quite different character depending on whether $\omega_{LO} > \omega_p$ or $\omega_p > \omega_{TO}$. The case $\omega_p \gg \omega_{LO}$ is discussed here. In semiconductors, the experimentalist may vary n_0, and hence ω_p, so the two situations $\omega_p < \omega_{LO}$ or $\omega_p > \omega_{LO}$ may be achieved in the same material.

The total dielectric function is denoted as $\varepsilon_{\text{total}}$. Perhaps this should just be called $\varepsilon(q, \omega)$. However, it is conventional to reserve the latter name for only the electron–electron parts of the dielectric function

$$\varepsilon(q, \omega) = 1 - v_q^{(\infty)} P(q, \omega) \tag{7.202}$$

In the case where $\omega_p > \omega_{LO}$, it is customary—although perhaps not necessary—to explicitly separate the electron–electron part of the effective interaction:

$$V_{\text{eff}} = \frac{v_q}{\varepsilon_{\text{total}}} = \frac{v_q^{(\infty)}}{\varepsilon(q, \omega)} + V_{\text{sc-ph}}(q, \omega) \tag{7.203}$$

The second term on the right is the screened electron–phonon interaction. It is defined by this equation, so $V_{\text{sc-ph}}$ is the difference between V_{eff} and the screened Coulomb interaction. This definition may be manipulated by combining these two terms:

$$V_{\text{sc-ph}} = \frac{v_q^{(\infty)} + V_{ph}}{1 - (v_q^{(\infty)} + V_{ph})P} - \frac{v_q^{(\infty)}}{1 - v_q^{(\infty)}P} \tag{7.204}$$

$$= \frac{V_{ph}}{\varepsilon[1 - (v_p^{(\infty)} + V_{ph})P]} \tag{7.205}$$

$$= \frac{V_{ph}}{\varepsilon^2[1 - V_{ph}P/\varepsilon]} \tag{7.206}$$

From the definition of $V_{ph} = M_q^2 D^{(0)}$, rewrite the effective interaction as

$$V_{\text{eff}} = \frac{v_q^{(\infty)}}{\varepsilon} + \bar{M}_q^2 D \tag{7.207}$$

$$\bar{M}_q = \frac{M_q}{\varepsilon} \tag{7.208}$$

$$D(q) = \frac{D^{(0)}}{1 - M_q^2 D^{(0)} P/\varepsilon} = \frac{2\omega_q}{\omega^2 - \omega_q^2 - 2\omega_q M_q^2 P(q)/\varepsilon} \tag{7.209}$$

The screened electron–phonon interaction $V_{\text{sc-ph}}$ is expressed as the product of the renormalized phonon Green's function $D(q, \omega)$ times the square of the screened matrix element $\bar{M}_q$. The phonon Green's function contains self-energy terms arising from the polarization of the electron gas. The quantity $P(q, \omega)$ describes the polarization of the electron gas, while $\bar{M}_q^2$ is the coupling between the electrons and phonons. It is important to note that this phonon Green's function is different from the one which would be derived from Dyson's equation by ignoring the electron–electron interactions. The latter result would be Dyson's equation with no electron–electron interactions:

$$D' = \frac{D^{(0)}}{1 - M_q^2 D^{(0)} P} \tag{7.210}$$

The difference is the additional factor of $1/\varepsilon$ in the self-energy term. The derivation of $D(q, \omega)$ includes electron–electron interactions properly and show that the factor of $1/\varepsilon$ actually belongs. Since ε is usually greater than unity, including the $1/\varepsilon$ factor will considerably weaken the self-energy effects.

Another interesting feature of the screened electron–phonon interaction is that the matrix element M_q is divided by the dielectric function ε. Since M_q is something in the nature of the electron–ion potential, it is reasonable to expect it to be divided by the dielectric function. The conduction electrons screen the electron–ion interaction. The ions are heavy and move with small frequencies compared to electronic response times. The electrons are able to follow the ion motion, which provides the screening.

There is a good reason for separating the effective interaction into the Coulomb and screened phonon parts. When $\omega_p \gg \omega_{LO}$, the dielectric function $\varepsilon(q, \omega)$ which occurs in the phonon part needs to be evaluated at $\omega \approx \omega_{LO}$. For $\omega_p \gg \omega_{LO}$, $\varepsilon(q, \omega_{LO})$ is well approximated

by the static limit of $\varepsilon(q)$. Then it is a good approximation to rewrite the effective interaction as

$$V_{\text{eff}} = \frac{v_q^{(\infty)}}{\varepsilon(q, \omega)} + \bar{M}_q^2 D(q, \omega) \tag{7.211}$$

$$\bar{M}_q = \frac{M_q}{\varepsilon(q)} \tag{7.212}$$

$$D(q) = \frac{2\omega_q}{\omega^2 - \omega_q^2 - 2\omega_q M_q^2 P(q)/\varepsilon(q)} \tag{7.213}$$

In evaluating the screened Coulomb term, the full frequency dependence must be retained in $\varepsilon(q, \omega)$. But the static limits $\varepsilon(q)$ and $P(q, \omega = 0)$ are satisfactory in the phonon term. Most calculations have used either Thomas-Fermi theory or the RPA for the dielectric function. Although it would be more accurate to use a Singwi–Sjölander dielectric function, the additional labor may not be worth the incremental increase in accuracy.

7.3.2. Experimental Verifications

First-order Raman scattering is permitted in crystals which lack an inversion center. Two examples are zincblende and wurtzite, which are the crystal structure of most III–V and II–VI semiconductors. The Raman experiment measures the frequency of excitations which have the same magnitude of wave vector as the incident light. For optical frequencies, the wave vector is typically 10^5 cm^{-1}, which is essentially the limit of $q \to 0$ for $P(q, \omega)$. In this limit, the electron-gas term $-v_q P(q, \omega)$ becomes $-\omega_p^2/\omega^2$, so the total dielectric function is

$$\lim_{q \to 0} \varepsilon_{\text{total}}(q, \omega) = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{1 - \omega^2/\omega_{TO}^2} - \omega_p^2/\omega^2 \tag{7.214}$$

The poles of $1/\varepsilon_{\text{total}}(\omega)$ occur where $\varepsilon_{\text{total}}(\omega) = 0$. Setting the above equation to zero gives a quadratic equation for ω^2, which is easily solved to give

$$\omega_\pm^2 = \tfrac{1}{2}\{\omega_p^2 + \omega_{LO}^2 \pm \sqrt{(\omega_p^2 + \omega_{LO}^2)^2 - 4\omega_p^2\omega_{TO}^2}\} \tag{7.215}$$

These two solutions are plotted as the solid lines in Fig. 7.9 as ω_p^2 is changed by increasing n_0. The parameters are appropriate for GaAs, which has $m^* = 0.072$, $\varepsilon_\infty = 11.3$, $\omega_{TO} = 268$ cm^{-1}, and $\omega_{LO} = 291$ cm^{-1}. Also shown is the experimental data of Mooradian and Wright (1966) for GaAs; they measured the frequencies by Raman scattering. There is excellent agreement between theory and experiment.

The LO phonons and the plasmons are two modes which are mutually coupled. Figure 7.9 represents a typical crossing phenomenon of two coupled modes. It is a particularly clear example, since both modes are completely classical. Even if $\hbar$ is restored to our equations, it would not appear in (7.215). It is also useful to understand the asymptotic limits of very high and low density. At high values of n_0, the analytical solution of (7.215) gives the roots when $\omega_p^2 \gg \omega_{LO}^2$

$$\omega_+^2 \to \omega_p^2 \tag{7.216}$$

$$\omega_-^2 \to \omega_{TO}^2 \tag{7.217}$$

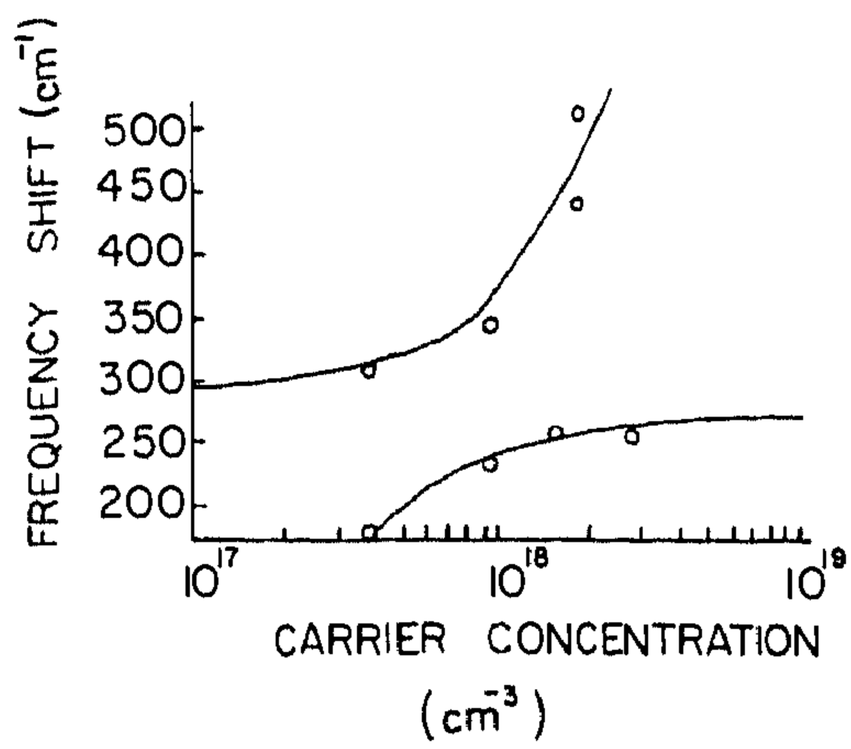


FIGURE 7.9 Plasma frequency in GaAs as a function of electron density n_0, as determined by Raman scattering. Solid line is theory using $\omega_{TO} = 268\ \text{cm}^{-1}$. $\omega_{LO} = 291\ \text{cm}^{-1}$, $\varepsilon_\infty = 11.3$, and $m^* = 0.07$. *Source*: Mooradian and Wright (1966) (used with permission).

The phonon-like mode has a frequency ω_{TO}, not ω_{LO}. The choice ω_{TO} occurs because the frequency difference between ω_{TO} and ω_{LO} is caused by long-ranged Coulomb interactions. The electron gas screens these long-ranged Coulomb effects whenever $\omega_p^2 \gg \omega_{LO}^2$. In this case, the electron gas can respond with a higher frequency than the phonons and can follow the motion of the phonons. The electron gas will screen the phonons and prevent long-range electric fields. The frequency difference between ω_{LO} and ω_{TO} vanishes at long wavelengths, and the longitudinal phonon excitation has frequency ω_{TO}. Similarly, the phonons cannot follow the plasma oscillations of the electron gas, so the phonons do not contribute to the screening of the electron–electron interactions. The plasma frequency in (7.201) has a dielectric constant ε_∞, not ε_0. The terms which contribute to ε_∞ are high-frequency interband transitions, which can certainly follow these low-frequency plasma oscillations.

The other limit of density has low values of n_0. When $\omega_p^2 \ll \omega_{LO}^2$, the two roots of (7.215) are

$$\omega_+^2 \to \omega_{LO}^2 \tag{7.218}$$

$$\omega_-^2 \to \frac{\omega_p^2 \omega_{TO}^2}{\omega_{LO}^2} = \frac{4\pi e^2 n_0}{\varepsilon_0 m^*} \tag{7.219}$$

The Lyddane–Sachs–Teller relationship has been used in the last identity. Now the phonon-like mode has a frequency ω_{LO}. The electron gas cannot oscillate as fast as the phonons and does not screen the long-ranged Coulomb fields of the phonons. The phonons have a longitudinal frequency given by ω_{LO}, which must be the correct limit as $n_0 \to 0$. Since the phonons can now follow the oscillations of the electron gas, they do contribute to the screening of the plasma oscillations. Hence the plasma frequency now contains the screening factor ε_0 rather than ε_∞.

7.3.3. Electron Self-Energies

The motions of the electrons are affected by the other electrons and by the phonons. These interactions determine the self-energy of the electron. The self-energy will have contributions from electron–electron as well as electron–phonon interactions. They combine into the screened interaction, described in Sec. 7.3.1 and summarized in (7.166) and (7.173). It is a rather cumbersome expression in its general form. Actual evaluation, to get analytical

results, will require making some approximations at some point in the calculation. The only calculations which have been reported, and verified experimentally, are those for the limit that $\omega_p \gg \omega_{LO}$. In this case it is permissible to separate the self-energy effects into an electron part and a screened phonon part. The electron–electron part may be evaluated by the procedures for the homogeneous electron gas given in Chapter 5. Nothing further needs to be mentioned about this term.

The present objective is to calculate the electron self-energy from electron–phonon interactions. The first step is to calculate the one-phonon self-energy term. For the screened electron–phonon interaction in (7.209), it is

$$\Sigma(p, ip) = -\frac{1}{\beta}\sum_{i\omega_n}\int\frac{d^3q}{(2\pi)^3}\frac{M_q^2}{\varepsilon(q, i\omega)^2}\mathscr{D}(\mathbf{q}, i\omega_n)\mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip + i\omega_n) \tag{7.220}$$

There should also be a summation over the different phonon modes λ, which is omitted, since the discussion will be confined to polar coupling. Later it will be decided whether this one-phonon term is sufficient. Perhaps more terms will be needed, with higher numbers of phonons. It will turn out that the one-phonon term is adequate, since the coupling is weak. The screening weakens the effective coupling constant. Materials which have intermediate values of α when they are insulators become weak coupling when the electron gas screens. The screening will be shown later in great detail. For the moment, the one-phonon self-energy is sufficient. Also defer, for the moment, the question of whether to use the Rayleigh–Schrödinger or Brillouin–Wigner form of the self-energy. The energy is designated as ip, which will be ξ_p for the Rayleigh–Schrödinger form of perturbation theory and ω for the other.

The one-phonon self-energy (7.220) will be evaluated for the degenerate semiconductor. A degenerate semiconductor has been doped with a sufficient number of carriers that they become an electron gas, with a well-defined Fermi surface, in the limit of zero temperature. The degenerate limit will simplify the calculation in a number of ways, since the homogeneous electron-gas results are used to describe the dielectric function $\varepsilon(q, i\omega)$.

The self-energy in (7.220) has a free-particle electron Green's function $\mathscr{G}^{(0)}$ but a renormalized phonon propagator $\mathscr{D}$. The procedure is to solve the phonon system assuming that the electrons are a free-particle gas. The self-energy terms calculated for the phonons have already been discussed. The phonons change their frequency at long wavelength, because of the screening of the electron gas. These newly found modes could be used in calculating the properties of the electrons; the phonon Green's function in (7.213) could be used in (7.220). Although this procedure is reasonable and systematic, it makes the analytical calculation too difficult. Instead, an unperturbed phonon Green's function $\mathscr{D}^{(0)}$ is used in (7.220). An Einstein model is assumed for the optical phonon frequencies:

$$\mathscr{D}^{(0)}(q, i\omega_n) = \mathscr{D}^{(0)}(i\omega_n) = \frac{-2\omega_0}{\omega_n^2 + \omega_0^2} \tag{7.221}$$

The basic problem with (7.220) is that there is too much physics in it. It describes how the electron gas screens and modifies the phonons and their interactions with electrons. It does this sufficiently well that it is too cumbersome to evaluate without making some approximations. One may as well make these approximations at the beginning of the calculation, since doing it early saves a lot of work later on. The first approximation has already been described, which is to replace the phonon Green's function $\mathscr{D}$ by the unperturbed propagator $\mathscr{D}^{(0)}$.

The next approximation is much more drastic. The dielectric function $\varepsilon(q, i\omega)$ will be approximated by its static value $\varepsilon(q)$. The frequency dependence of the dielectric function is ignored. This approximation is reasonable as long as the phonon frequencies are much less than the plasma frequency ω_p. The approximation $\varepsilon(q, \omega) \approx \varepsilon(q)$ is allowable whenever $\omega_{LO} \ll \omega_p$. In fact, the dielectric function may now be combined with the matrix element to give a screened interaction $V_s(q) = M_q^2/\varepsilon(q)^2$. The self-energy function in (7.220) now has the form

$$\Sigma(p, ip) = -\frac{1}{\beta}\sum_{i\omega_n}\int\frac{d^3q}{(2\pi)^3}V_s(q)\mathscr{D}^{(0)}(i\omega_n)\mathscr{G}^{(0)}(\mathbf{p}+\mathbf{q}, ip+i\omega_n) \tag{7.222}$$

The summation over Matsubara frequencies now has the simple form which was evaluated in Sec. 3.5:

$$\Sigma(p, ip) = \int\frac{d^3q}{(2\pi)^3}V_s(q)\left[\frac{1+N_0-n_F(\xi_{\mathbf{p}+\mathbf{q}})}{ip-\xi_{\mathbf{p}+\mathbf{q}}-\omega_0}+\frac{N_0 n_F(\xi_{\mathbf{p}+\mathbf{q}})}{ip-\xi_{\mathbf{p}+\mathbf{q}}+\omega_0}\right] \tag{7.223}$$

A much more complicated result is obtained when the more accurate forms of $\varepsilon(q, i\omega)$ and $\mathscr{D}(q, i\omega)$ are retained in doing the summations. The simple result (7.223) is the basis for our discussion.

The previous calculations of polaron theory always occurred in one-electron systems in which the Fermi occupation factors n_F could be set equal to zero. In the present discussion, they must be retained, and they play an important role in the nature of the final result. Only the case for the zero temperature is discussed, where the phonon thermal occupation factors $N_0 = 1/(e^{\beta\omega_0}-1)$ can be set equal to zero. Next consider the wave vector integrals, where the integration variables are changed: $\mathbf{k} = \mathbf{p}+\mathbf{q}$

$$\Sigma(p, ip) = \int\frac{d^3k}{(2\pi)^3}V_s(\mathbf{p}-\mathbf{k})\left[\frac{1-n_F(\xi_{\mathbf{k}})}{ip-\xi_{\mathbf{k}}-\omega_0}+\frac{n_F(\xi_{\mathbf{k}})}{ip-\xi_{\mathbf{k}}+\omega_0}\right] \tag{7.224}$$

A model form must be chosen for the dielectric function $\varepsilon(q)$. Of course, it would be super-accurate to use the Singwi–Sjölander form, but it is sufficiently complicated that a computer is needed to find any results. The Thomas–Fermi form is used instead. Although less accurate, it permits an analytical answer. Later it is argued that the answer is insensitive to the screening, so that the Thomas–Fermi approximation is probably adequate. However, this assertion has never been thoroughly checked by doing the self-energy calculation with the several alternate dielectric models of Chapter 5.

The retarded self-energy has both real and imaginary parts. The imaginary part will be done first. It can be evaluated exactly now that enough approximations have been made. It is given by ($ip \to \omega + i\delta$)

$$\begin{aligned}\mathrm{Im}[\Sigma(p, \omega)] = -\frac{\pi}{(2\pi)^3}\int k^2dk\{&[1-n_F(\xi_k)]\delta(\omega-\xi_k-\omega_0)\\ &+n_F(\xi_k)\delta(\omega-\xi_k+\omega_0)\}\int d\Omega_k V_s(\mathbf{p}-\mathbf{k})\end{aligned} \tag{7.225}$$

The delta functions take out the dk integrals. There remains just the angular integrals. Let $\nu = \cos\theta$ denote the angle between $\mathbf{p}$ and $\mathbf{k}$. For polar coupling, with Thomas–Fermi screening, the angular integrals are

$$\int d\Omega_k V_s(\mathbf{p}-\mathbf{k}) = \frac{8\pi^2\alpha\omega_0^{3/2}}{\sqrt{2m}}\int_{-1}^{1}d\nu\frac{(\mathbf{p}-\mathbf{k})^2}{[q_{TF}^2+(\mathbf{p}-\mathbf{k})^2]^2} \tag{7.226}$$

The integrals are straightforward and give a lengthy result:

$$\mathrm{Im}[\Sigma(p,\omega)] = -\frac{\alpha\omega_0^{3/2}}{4\sqrt{\varepsilon_p}}[\Theta(\omega-\omega_0)g_P(\omega-\omega_0)+\Theta(-\omega-\omega_0)g_P(\omega+\omega_0)]$$

$$g_p(Z) = \varepsilon_s\left\{\frac{1}{\varepsilon_s+[\sqrt{\varepsilon_p}+\sqrt{Z+\varepsilon_p}]^2}-\frac{1}{\varepsilon_s+[\sqrt{\varepsilon_p}-\sqrt{Z+\varepsilon_p}]^2}\right\}$$
$$+\ln\left[\frac{\varepsilon_s+[\sqrt{\varepsilon_p}+\sqrt{Z+\varepsilon_p}}{\varepsilon_s+[\sqrt{\varepsilon_p}-\sqrt{Z+\varepsilon_p}}\right] \tag{7.227}$$

$$\varepsilon_s = \frac{\hbar^2 q_{TF}^2}{2m} \tag{7.228}$$

The imaginary self-energy is shown in Fig. 7.10. It is plotted as a function of the energy ω relative to the Fermi energy for some value of p. The striking feature in the result is that $\mathrm{Im}(\Sigma)$ is zero for energies within ω_0 of the Fermi surface. Those electrons outside of this range can lose energy by emitting phonons. But those electrons within ω_0 of the Fermi energy cannot emit phonons and have a zero value for $\mathrm{Im}(\Sigma)$.

The gap in the allowed values of $\mathrm{Im}(\Sigma)$ can be understood in a simple way. The excitations with energy $\omega > \omega_0$ above the Fermi sea are electrons, or electron-like quasiparticles. They can decay by emitting an LO phonon of energy ω_0. If they initially have an energy ω, then after emitting the optical phonon, they have a final energy $\omega - \omega_0$. The electron states up to the energy E_F, or ($\omega = 0$), are all occupied, so that the final state must have an energy larger than this value. Otherwise the transition cannot occur since the final states are all occupied. This reasoning explains why $\omega - \omega_0 > 0$ for electron emission of a phonon, and why $\mathrm{Im}(\Sigma) = 0$ for $0 < \omega < \omega_0$. Phonon emission by electrons can occur only for initial energies at least ω_0 above the Fermi energy. These processes are illustrated qualitatively in Fig. 7.11(a). The allowed transition has enough initial energy that the final state is above the occupied levels. The forbidden transition would go to a state already occupied.

The behavior for $\omega < 0$ may be understood by realizing that the excitations are not electrons. Instead, the excitations below the Fermi energy are states where an electron is missing. These are called *Holes* in a metal. Unfortunately, the word *hole* in a semiconductor usually means an excitation of the valence band. Our word Hole, with a capital "H," is an excitation of the conduction band. It has an electron missing from the otherwise filled Fermi sea. These Holes decay by having an electron jump into this empty state, which is illustrated in Fig. 7.11(b). The decay of a Hole is allowed whenever the Hole can be filled by an electron already present. If the Hole has initial energy $\omega < -\omega_0$, then an electron in the Fermi sea can

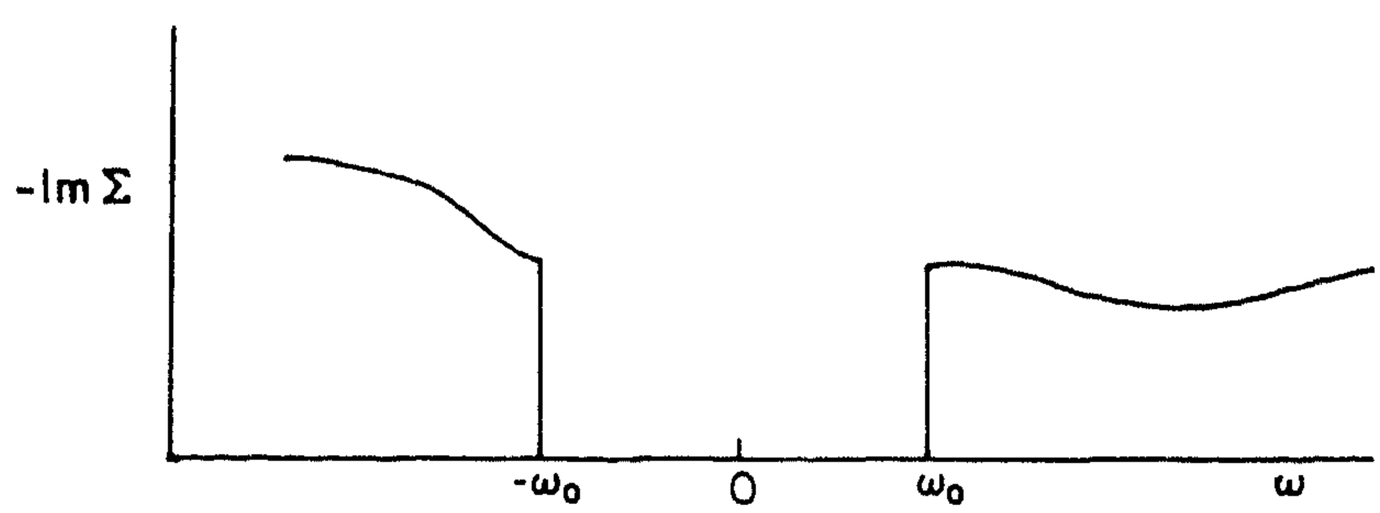


FIGURE 7.10 The imaginary part of the electron self-energy from scattering by optical phonons. The energy ω is measured from the Fermi surface.

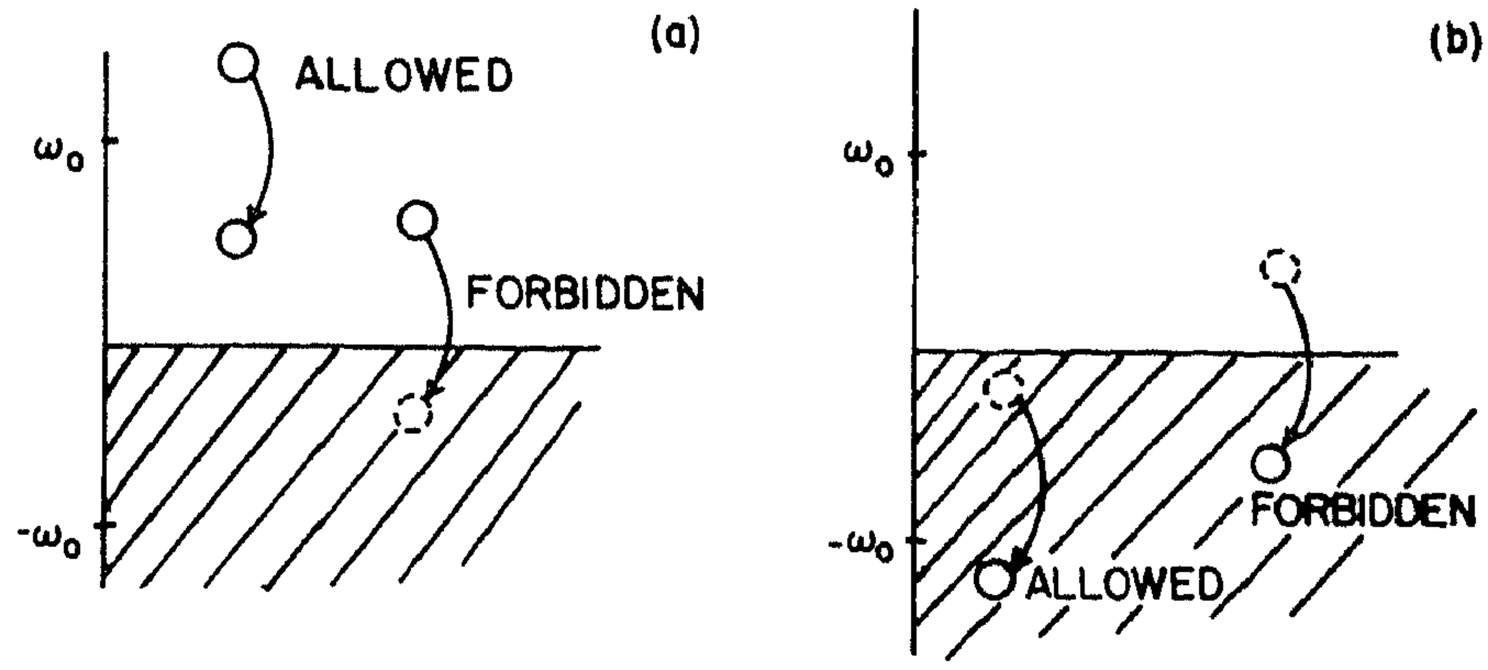


FIGURE 7.11 The explanation of why $\mathrm{Im}(\Sigma) = 0$ for excitation energies within ω_0 of the Fermi surface. (a) Electrons with $0 < \omega < \omega_0$ cannot decay since the final state is occupied (this transition is listed as forbidden). (b) For $-\omega_0 < \omega < 0$, hole states cannot decay since no electrons are available.

emit an optical phonon and jump into this Hole state. This process is shown as the "allowed" transition in Fig. 7.11(b). The process is equivalent to the Hole jumping to lower energy by an amount ω_0 since the zero energy for a Hole is the Fermi energy. The "forbidden" process of Hole decay in Fig. 7.11 cannot occur because it requires the initial electron to be above the Fermi sea. It is not possible at zero temperature if the system is in equilibrium. The forbidden hole decay explains why $\mathrm{Im}(\Sigma)$ is zero for $-\omega_0 < \omega < 0$.

These arguments explain the energy gap in $\mathrm{Im}(\Sigma)$ for $-\omega_0 < \omega < \omega_0$. Of course, this gap exists only for the self-energy contribution from optical phonons. The electron will have nonzero values of $\mathrm{Im}(\Sigma)$ in this energy region arising from scattering by acoustical phonons and from electron–electron interactions.

The self-energies describe the excitation spectrum of the interacting system. They are electron-like quasiparticles for energies above the Fermi energy and Holes below the Fermi energy.

The function $g_p(Z)$ in (7.227) is a smooth function of both p and Z. This lack of structure is due to the screening of the electron gas, which eliminates all sharp features. Since $g_p(Z)$ is sufficiently dull, it is a good approximation to treat it as a constant. A simple place to evaluate this constant is at $Z = 0$ and $p = 0$. This limit defines a coupling constant called f_0:

$$f_0 = \lim_{p\to 0}\lim_{Z\to 0}\frac{\alpha\omega_0^{3/2}}{4\sqrt{\varepsilon_p}}g_p(Z) = \frac{\alpha(\omega_0 E_F)^{3/2}}{(E_F+\varepsilon_s)^2} \tag{7.229}$$

$$\mathrm{Im}[\Sigma(\omega)] = -f_0[\Theta(\omega-\omega_0) - \Theta(-\omega-\omega_0)] \tag{7.230}$$

The real part of the self-energy may now be obtained by a simple argument. The retarded self-energy function $\Sigma(p, \omega)$ is causal and hence obeys a Kramers–Kronig relation (see Sec. 5.7). The real part may be derived from the imaginary part by

$$\mathrm{Re}[\Sigma(\omega)] = \int_{-\infty}^{\infty}\frac{d\omega'}{\pi}\frac{\mathrm{Im}[\Sigma(\omega')]}{\omega'-\omega} \tag{7.231}$$

$$= -\frac{f_0}{\pi}\left(\int_{-\infty}^{-\omega_0}\frac{d\omega'}{\omega'-\omega} + \int_{\omega_0}^{\infty}\frac{d\omega'}{\omega'-\omega}\right) \tag{7.232}$$

$$= -\frac{f_0}{\pi}\ln\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right| \tag{7.233}$$

The result (7.233) was first obtained by Englesberg and Schrieffer (1963). They were actually discussing metals with optical phonons, but the result also applies to doped semiconductors. The total effect of the screened polar interaction is contained in the coupling constant f_0. The same type of answer is obtained for any electron gas with an Einstein phonon, but the detailed definition of f_0 is altered. A consequence of electron screening is that the details of the coupling constant are not terribly important.

There have been several attempts to verify this theory, in degenerate semiconductors, through electron tunneling experiments. Usually the tunneling is done through the Schottky barrier at a metal–semiconductor interface. The metal effects are negligible, and all structure in the electron tunneling conductance can be attributed to the semiconductor unless the metal is superconducting, and then the Bardeen–Cooper–Schrieffer (BCS) gap is apparent. But for normal metals, even if ferromagnetic, the metal–semiconductor junction is determined by the properties of the semiconductor (Conley and Mahan, 1967).

The most direct confirmation of the theory was obtained by Tsui (1974), who measured the constant f_0 in InAs as the magnitude of the step increase in the scattering rate of the electrons. He was measuring metal–semiconductor junctions in a magnetic field. The differential conductance d^2I/d^2V showed oscillations, similar to the deHaas–van Alphen

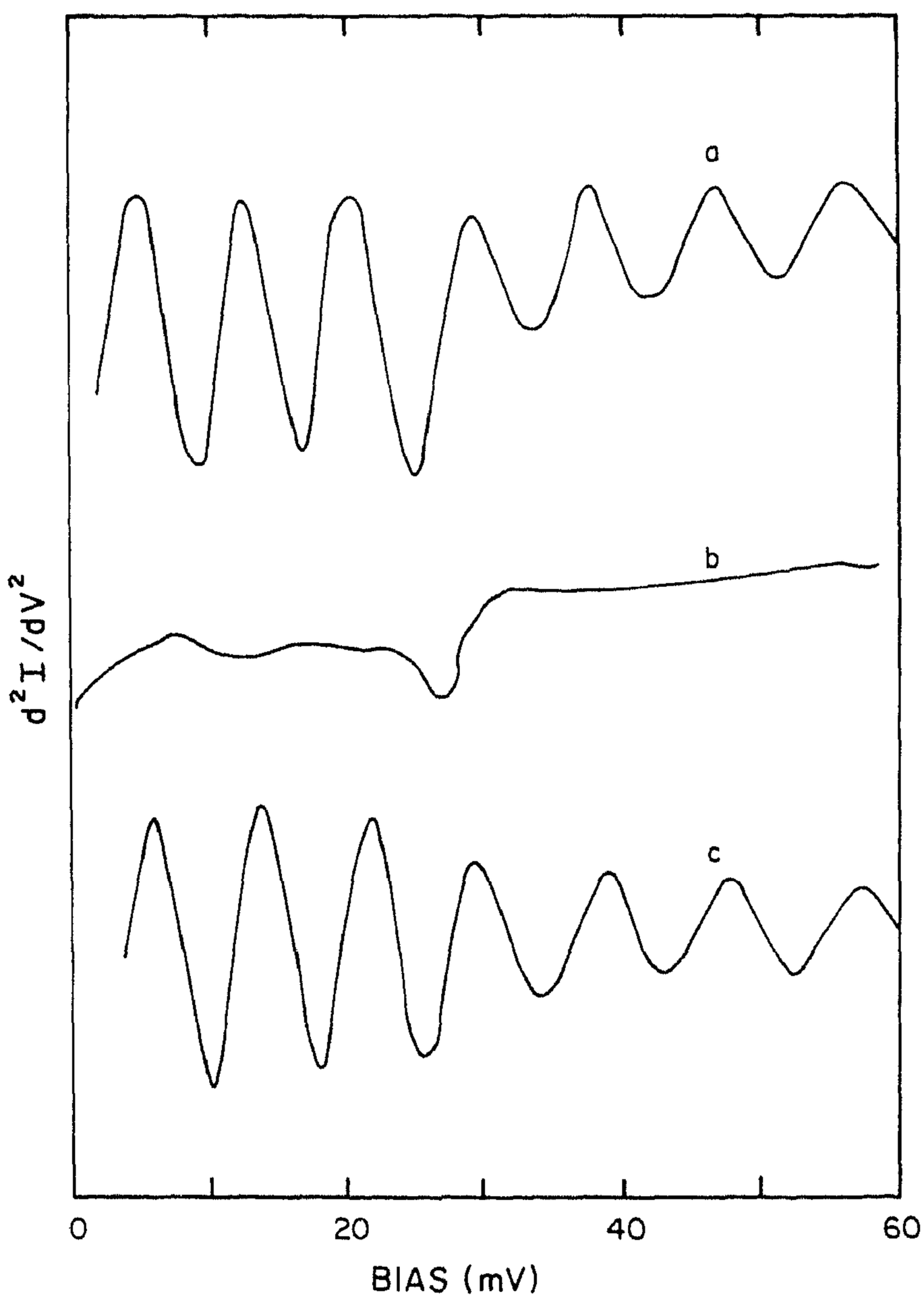


FIGURE 7.12 d^2I/dV^2 vs V data from an InAs–oxide–Pb junction. The InAs sample has $n_0 = 5.5 \times 10^{17}$ cm^{-3} and $T = 4.2$ K. (a) Taken with $H = 22$ kG, (b) taken with $H = 2$ kG, and (c) the difference between (a) and (b). *Source*: Tsui (1974) (used with permission).

(dHvA) oscillations in a metal. Tsui's data is shown in Fig. 7.12. The oscillations change their amplitude at voltages greater than 30 meV, which is the energy of the LO phonon in InAs. The tunneling voltage is the difference of the two chemical potentials and is the energy in the semiconductor which the electrons have after they tunnel from the metal. Using parameters for InAs, Tsui used (7.229) to predict a value for $\tau_0 = \hbar/(2f_0)$ and compared it with his experimental number:

$$\text{Theory:} \qquad \tau_0 = 5.3 \times 10^{-13}\ \text{s} \tag{7.234}$$

$$\text{Experiment:} \qquad \tau_0 = (5.1 \pm 0.3) \times 10^{-13}\ \text{s} \tag{7.235}$$

There is excellent agreement between theory and experiment.

In many respects, the degenerate semiconductor is the ideal environment in which to study the electron gas and polaron effects in an electron gas. The electron gas can be made to have an effective high density at which the theory is supposed to be valid. All parameters of the theory, such as α or m^*, can be determined by independent experiments in the insulating semiconductor. Theory can be done with no adjustable parameters. The experimentalists have systems which are relatively easy to measure and which have an electron density which may be varied continuously.

7.4. METALS

Electron–phonon interactions in metals have been studied for several reasons. The phonons in metals are significantly affected by the electron gas, so any study of the phonon system requires an understanding of the electron–phonon interaction. The electrons near the Fermi surface have their motions influenced by the phonons, so that many experiments which measure these electrons, such as transport or cyclotron resonance, are affected by electron–phonon interactions. The interaction also causes superconductivity in many metals. There are many experimental methods which have been applied to this topic, and ample reasons to do so.

There is an initial problem facing any theoretical treatment. Since the two systems of electrons and phonons affect each other very much, it is not obvious where to begin the many-body calculations. Does one first solve the phonons assuming the electrons have no phonon influences, or solve for the electron properties using the free propagators for phonons? That is, does one first solve for the phonon or the electron self-energies? Neither approach may work if the two systems are strongly coupled. Actually, there is a right way to begin, as discussed by Migdal (1958). The justification of the procedure requires, to some extent, knowing the solution to the problem. One should start by first discussing the phonons using electron states which do not contain phonon effects. Although the phonons affect the electrons, they alter electron states only within a Debye energy of the Fermi surface. These states are only a small fraction of the electrons in the system. However, in solving for the electrons' influence upon the phonons, the average is taken over all the occupied states of the electron gas. This average is influenced to a negligible fashion by the few electrons at the Fermi surface, which themselves are influenced by phonons. A good answer is obtained by first solving for the phonon states using electron states which have no phonons included. Then one can solve for the electron states using these renormalized phonon states. The standard procedure is to first solve for the properties of the phonons. An electron Green's function is employed which does not have a self-energy contribution from phonons. But it does contain

the effects of electron–electron interactions, which are very important for obtaining the correct physical description.

7.4.1. Phonons in Metals

For simplicity, the discussion is limited to metals, such as sodium or aluminum, which have only one atom per unit cell. The many-body system is a collection of ions and electrons. It is charge neutral, so the ions and electrons have the same average charge density $|e|n_0 = eZn_i$, where n_0 and n_i are the particle densities of electrons and ions. The ions of valence Z are considered as rigid objects. The excitations of the inner core electrons can be neglected, since these energetic events are not induced by the motion of the ions. The core excitations will contribute a small amount of dielectric screening, which can be included as a phenomenological high-frequency dielectric constant ε_i. The Hamiltonian was discussed earlier in Sec. 1.3:

$$H = \sum_i \frac{p_i^2}{2m_i} + \sum_\alpha \frac{P_\alpha^2}{2M_\alpha} + \frac{e^2}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i\alpha} V_{ei}(\mathbf{r}_i - \mathbf{R}_\alpha) + \tfrac{1}{2}\sum_{\alpha\beta} V_{ii}(\mathbf{R}_\alpha - \mathbf{R}_\beta) \tag{7.236}$$

The potential $V_{ei}(\mathbf{r}_i - \mathbf{R}_\alpha)$ is between an electron at $\mathbf{r}_i$ and an ion at $\mathbf{R}_\alpha$. It was called V_A in Sec. 1.3. Similarly, the potential $V_{ii}(\mathbf{R}_\alpha - \mathbf{R}_\beta)$ is between two ions. Both of these potentials are unscreened, so they behave at large distances as

$$\lim_{r\to\infty} V_{ei}(r) = -\frac{Ze^2}{\varepsilon_i r} \tag{7.237}$$

$$\lim_{R\to\infty} V_{ii}(R) = -\frac{Z^2e^2}{\varepsilon_i R} \tag{7.238}$$

In a free-electron metal such as sodium or aluminum, the ions are sufficiently far apart that the Coulomb form (7.238) of V_{ii} is probably valid for all ion pairs. However, the electron–ion interaction V_{ei} must always be treated realistically at small distances. The customary procedure is to use a pseudopotential (Harrison, 1966). The electron–electron interactions are also included in the Hamiltonian. They cause the potentials to be screened.

The first discussion is at an introductory level, approximately the same as Schrieffer (1964). In this model the ions are point charges, and the electron gas is jellium. Later real metal effects are discussed at a more advanced level. The introductory treatment will try to explain the physics and avoid the bewildering notation of the correct treatment.

The first step is to solve the Hamiltonian of the electrons in the absence of phonons. The ions are fixed in their equilibrium positions $\mathbf{R}_\alpha^{(0)}$. Then (7.236) becomes the Hamiltonian of electrons in a periodic, neutral system:

$$H_{0e} = \sum_i \frac{p_i^2}{2m_i} + \frac{e^2}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i\alpha} V_{ei}(\mathbf{r}_i - \mathbf{R}_\alpha^{(0)}) + \tfrac{1}{2}\sum_{\alpha\beta} V_{ii}(\mathbf{R}_\alpha^{(0)} - \mathbf{R}_\beta^{(0)}) \tag{7.239}$$

The solution of this problem is quite formidable. The electrons states are Bloch waves in the periodic potential. They must be calculated with the full effects of correlation and exchange in the metallic environment, which has only been done approximately. The problem is quite difficult, but it is not the problem at hand. For the discussion of electron–phonon effects, it is assumed that the electron part of the Hamiltonian has been solved. Furthermore, in this

introductory treatment, the electron states of our quasi-jellium model of a solid are approximated as plane waves.

The Hamiltonian of the phonons is first solved without reference to the electrons. In the harmonic approximation (see Sec. 1.1) the ions are assumed to have small displacements $\mathbf{Q}_\alpha$ about their equilibrium positions $\mathbf{R}_\alpha^{(0)}$, so the Hamiltonian can be written as

$$\mathbf{R}_\alpha = \mathbf{R}_\alpha^{(0)} + \mathbf{Q}_\alpha \tag{7.240}$$

$$H = H_{0e} + H_{0p} + H_{ep} \tag{7.241}$$

$$H_{0p} = \sum_\alpha \frac{P_\alpha^2}{2M_\alpha} + \tfrac{1}{4}\sum_{\alpha\beta}(\mathbf{Q}_\alpha - \mathbf{Q}_\beta)_\mu(\mathbf{Q}_\alpha - \mathbf{Q}_\beta)_\nu \Phi_{\mu\nu}(\mathbf{R}_\alpha^{(0)} - \mathbf{R}_\beta^{(0)}) \tag{7.242}$$

$$H_{ep} = \sum_{j\alpha} \mathbf{Q}_\alpha \cdot \nabla V_{ei}(\mathbf{r}_j - \mathbf{R}_\alpha^{(0)}) \tag{7.243}$$

$$\Phi_{\mu\nu}(\mathbf{R}) = \nabla_\mu \nabla_\nu V_{ii}(\mathbf{R}) \tag{7.244}$$

The term H_{0p} is called the bare-phonon Hamiltonian. These phonons are calculated by ignoring the electron–electron and electron–phonon interactions. The ion motions are calculated using only the direct ion–ion interaction potential V_{ii}. The bare-phonon Hamiltonian, once solved, describes a set of eigenfrequencies and normal modes which are quite unrealistic. The poor results are obtained because the phonon Hamiltonian is for a system which is not charge neutral. When the ions vibrate, the bare-phonon Hamiltonian H_{0p} treats the electron gas as a rigid background, which does not follow the ion motion. Hence the ion vibrations cause long-range Coulomb fields. These vibrations are just plasma oscillations, of the ions, so the long-wavelength phonons will have the frequency

$$\omega_{ip}^2 = \frac{4\pi Z^2 e^2 n_i}{M\varepsilon_i} \tag{7.245}$$

In usual solids, and metals, the long-wavelength oscillations of the atoms or ions are described by acoustical phonons, which have the dispersion relation $\omega(q) = C_s q$ proportional to the wave vector, where C_s is the speed of sound. In solids with more than one atom per unit cell there may also be optical phonons, which have a constant frequency at long wavelength. However, the above plasma modes of H_{0p} are not from optical modes of vibration, since the present theory is for solids with only one atom per cell and no optical modes. In fact, the bare-phonon Hamiltonian H_{0p} is sufficiently unrealistic that it does not have acoustical mode solutions, which are necessary for real metals. The ion plasma modes are the correct long-wavelength normal modes of H_{0p} but do not realistically approximate the actual modes in a metal.

In normal metals, the electrons and ions respond differently because of the great difference in their masses. When the ions vibrate, the electrons follow the ion motion. There are no long-range Coulomb fields and no ion plasma waves. When the electrons oscillate at their natural frequency ω_p, the ions are too heavy to follow, so the electron motion is not screened. The electrons can freely oscillate at their plasma frequency, while the ions cannot. Our bare-phonon Hamiltonian gives the wrong answer because it did not let the electrons follow the ion motion. Permitting the electrons to follow the ion motion is another set of words to describe the screening by the electrons of the motion of the ions.

The method of solving (7.242) for the bare phonons was described in Sec. 1.1. The ion displacements $\mathbf{Q}_\alpha$ and conjugate momenta $\mathbf{P}_\alpha$ are expanded in a set of normal modes,

$$\mathbf{Q}_\alpha = \frac{1}{\sqrt{N_i}} \sum_{\mathbf{k}} \mathbf{Q}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_\alpha} \tag{7.246}$$

$$\mathbf{P}_\alpha = \frac{1}{\sqrt{N_i}} \sum_{\mathbf{k}} \mathbf{P}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_\alpha} \tag{7.247}$$

$$H_{0p} = \sum_{\mathbf{k}} \left[\frac{1}{2M} \mathbf{P}_k \cdot \mathbf{P}_{-k} + \tfrac{1}{2} \mathbf{Q}_{\mathbf{k}\mu} \mathbf{Q}_{-\mathbf{k}\nu} \phi_{\mu\nu}(\mathbf{k}) \right] \tag{7.248}$$

$$\phi_{\mu\nu}(\mathbf{k}) = \bar{\Phi}(0) - \bar{\Phi}(\mathbf{k}) \tag{7.249}$$

$$\bar{\Phi}(\mathbf{k}) = \sum_{\alpha} e^{i\mathbf{k}\cdot\mathbf{R}_\alpha^{(0)}} \Phi_{\mu\nu}(\mathbf{R}_\alpha^{(0)}) \tag{7.250}$$

If the ions were point charges, these terms would be

$$V_{ii}(R) = \frac{Z^2 e^2}{\varepsilon_i R} \tag{7.251}$$

$$\Phi_{\mu\nu}(\mathbf{R}) = -\frac{Z^2 e^2}{\varepsilon_i} \left(\frac{\delta_{\mu\nu}}{R^3} - \frac{3R_\mu R_\nu}{R^5} \right) \tag{7.252}$$

$$\bar{\Phi}_{\mu\nu}(\mathbf{k}) = -\frac{4\pi Z^2 e^2}{\varepsilon_i} \sum_{\mathbf{G}} \frac{(k_\mu + G_\mu)(k_\nu + G_\nu)}{(\mathbf{k} + \mathbf{G})^2} \tag{7.253}$$

where $\mathbf{G}$ are the reciprocal lattice vectors for the crystal. The summation over $\mathbf{G}$ converges slowly, which may be helped by using Ewald techniques. The normal modes of the bare-phonon system are found as the eigenfrequencies $\Omega_{\mathbf{q}\lambda}$ and eigenstates $\hat{\xi}_{\mathbf{q}\lambda}$ of the equation

$$\det|M\Omega_{\mathbf{q}\lambda}^2 \delta_{\mu\nu} - \phi_{\mu\nu}(\mathbf{q})| = 0 \tag{7.254}$$

These frequencies and eigenstates are used to define a set of creation and destruction operators through the expansion in Eq. (1.85):

$$\mathbf{Q}_{\mathbf{q}} = \sum_{\lambda} \left(\frac{\hbar}{2\rho v \Omega_{\mathbf{q}\lambda}} \right)^{1/2} \hat{\xi}_{\mathbf{q}\lambda} A_{\mathbf{q}\lambda} \tag{7.255}$$

$$A_{\mathbf{q}\lambda} = a_{\mathbf{q}\lambda} + a_{-\mathbf{q}\lambda}^\dagger \tag{7.256}$$

$$H_{0p} = \sum_{\mathbf{q}\lambda} \Omega_{\mathbf{q}\lambda} (a_{\mathbf{q}\lambda}^\dagger a_{\mathbf{q}\lambda} + \tfrac{1}{2}) \tag{7.257}$$

where ρ is the density. The formal solution of the bare-phonon system is complete. The modes which are obtained are conceptually useful, but there is no need to calculate them. The set of eigenfrequencies and eigenstates are only used as the basis for solving the electron–ion interaction H_{ep}. When solving the interacting system, a new equation is obtained for the phonon modes. These new equations contain electron screening and provide a realistic description of the atomic motions. These are worth solving numerically. The eigenstates of the

bare-phonon system are useful only as a concept upon which to build our solutions in terms of Green's functions. The interaction term between the electron and phonons is written as

$$H_{ep} = \frac{1}{\sqrt{\nu}} \sum_{\mathbf{q}\lambda\mathbf{G}} M_\lambda(\mathbf{q}+\mathbf{G}) e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} A_{\mathbf{q}\lambda} \tag{7.258}$$

$$M_\lambda(\mathbf{q}+\mathbf{G}) = \left(\frac{\hbar}{2\rho\Omega_{\mathbf{q}\lambda}}\right)^{1/2} \hat{\xi}_{\mathbf{q}\lambda} \cdot (\mathbf{q}+\mathbf{G}) V_{ei}(\mathbf{q}+\mathbf{G}) \tag{7.259}$$

The phonon states in this basis are the bare phonons. It is an unrealistic basis but serves only as the starting point for the Green's function calculation. The electron states are put into the second quantized notation, so that the effective Hamiltonian has the form

$$\begin{aligned} H = &\sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \sum_{\mathbf{q}\lambda} \Omega_{\mathbf{q}\lambda} a^\dagger_{\mathbf{q}\lambda} a_{\mathbf{q}\lambda} + \frac{1}{2\nu} \sum_{\mathbf{pkq}} \sum_{ss'} v_q C^\dagger_{\mathbf{k}+\mathbf{q},s} C^\dagger_{\mathbf{p}-\mathbf{q},s'} C_{\mathbf{p}s'} C_{\mathbf{k},s} \\ &+ \frac{1}{\sqrt{\nu}} \sum_{\mathbf{q}\lambda\mathbf{p}\sigma} M_\lambda(\mathbf{q}) C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} A_{\mathbf{q}\lambda} \end{aligned} \tag{7.260}$$

The solution to this Hamiltonian was discussed in Sec. 7.3. The present problem corresponds to the case where the electron plasma frequency ω_p is very much larger than the phonon frequencies. In this case the effective interaction between two electrons may be written as (7.209) with a screened Coulomb interaction and a screened phonon interaction:

$$V_{\text{eff}}(\mathbf{q}, i\omega) = \frac{v_q}{\varepsilon(\mathbf{q}, i\omega)} + \sum_\lambda \bar{M}^2_\lambda(\mathbf{q}) \mathscr{D}_\lambda(\mathbf{q}, i\omega) \tag{7.261}$$

$$\varepsilon(\mathbf{q}, i\omega) = 1 - \frac{v_q}{\varepsilon_i} P(\mathbf{q}, i\omega) \tag{7.262}$$

$$\mathscr{D}_\lambda(\mathbf{q}, i\omega) = \frac{\mathscr{D}^{(0)}_\lambda}{1 - M^2_\lambda \mathscr{D}^{(0)}_\lambda P/\varepsilon} \tag{7.263}$$

where $\bar{M}_\lambda$ is the screened matrix element $\bar{M} = M/\varepsilon$. The quantity $\mathscr{D}_\lambda$ is the phonon Green's function. Its denominator contains the factor $M^2_\lambda \mathscr{D}^{(0)}_\lambda P/\varepsilon$. The term $1/\varepsilon$ comes from the screening of the electron gas. It describes how the electrons can follow the motion of the ions and hence screen out the long-range Coulomb forces. The screening eliminates the ion-plasma solutions and makes the modes acoustical. To show this, first rewrite the phonon Green's function as

$$D_{\text{ret}}(\mathbf{q}, \omega) = \frac{2\Omega_{\mathbf{q}\lambda}}{\omega^2 - \Omega^2_{\mathbf{q}\lambda} - 2\Omega_{\mathbf{q}\lambda}(M^2_\lambda P/\varepsilon)} \tag{7.264}$$

At long wavelength, the factors in the denominator of this Green's function have the following limits:

$$\lim_{q\to 0} \begin{cases} \Omega_{\mathbf{q}\lambda} \to \omega_{ip} \\ M_\lambda \to q\left(\dfrac{\hbar}{2\rho\omega_{ip}}\right)^{1/2} \dfrac{4\pi Z e^2}{q^2 \varepsilon_i \nu_0} \\ 2\Omega_{\mathbf{q}\lambda} M^2_\lambda \to \dfrac{v_q}{\varepsilon_i} \omega^2_{ip} \\ \omega^2 - \Omega^2_{\mathbf{q}\lambda} - 2\Omega_{\mathbf{q}\lambda}(M^2 P/\varepsilon) \to \omega^2 - \omega^2_{ip} - \omega^2_{ip}\left(\dfrac{1}{\varepsilon} - 1\right) = \omega^2 - \omega^2_{ip}/\varepsilon \end{cases} \tag{7.265}$$

The last term on the right shows that the phonon energy denominator becomes the simple expression $\omega^2 - \omega_{ip}^2/\varepsilon$. For the small frequencies of phonons, the static dielectric function $\varepsilon(q, 0)$ may be used. In the limit $q \to 0$ it is given by the compressibility sum rule (5.266). The zero of the phonon energy denominator defines the new phonon modes $\omega_\lambda^2(q) = \omega_{ip}^2/\varepsilon(q)$

$$\lim_{q\to 0} \omega_\lambda(q) = qC_s \tag{7.266}$$

$$C_s = \frac{\omega_{ip}}{q_{TF}}\sqrt{\frac{K_f}{K}} \tag{7.267}$$

The long-wavelength excitation of the coupled system of ions and electrons are now acoustical phonons. The speed of sound is given by C_s. The formula for C_s is similar to the result first obtained by Bohm and Staver (1951), who found that $C_s = \omega_{ip}/q_{TF} = v_F\sqrt{Zm/3M}$. These ordinary sound waves are obtained because the electron gas follows the vibrations of the ions. When the ion of charge Ze moves, an amount of screening charge nearly $-Ze$ follows it, so the vibrating entity is nearly neutral. No long-range Coulomb fields are present, and there are no waves at the plasma frequency of the ions. The screening charge has contributions from all electrons in the Fermi sea.

The situations in metals is quite analogous to that of semiconductors. There the electron–electron interactions caused the LO phonon modes to be given by

$$\omega^2 = \omega_{TO}^2 + \frac{\omega_{LO}^2 - \omega_{TO}^2}{\varepsilon(q)} \tag{7.268}$$

To apply this formula to metals, set $\omega_{LO} = \omega_{ip}$ and $\omega_{TO} = 0$. The TO frequency is zero in metals since there is no restoring force for transverse oscillations. These steps give the same formula $\omega^2 = \omega_{ip}^2/\varepsilon(q)$.

The phonon Green's function can be written approximately as

$$D_\lambda(\mathbf{q}, \omega) = \frac{2\Omega_{\mathbf{q}\lambda}}{\omega^2 - \omega_\lambda^2(\mathbf{q})} \tag{7.269}$$

This approximation is valid only when the damping of the phonon states can be ignored. The phonon eigenfrequencies $\omega_\lambda(\mathbf{q})$ are the ones found with the inclusion of electron–electron interactions and represent the actual phonon modes measured in the solid. The electron–phonon matrix element $M_{\mathbf{q}\lambda}^2$ in (7.258) contains the factor $\Omega_{\mathbf{q}\lambda}^{-1}$. In the screened phonon interaction, the unphysical frequency $\Omega_{\mathbf{q}l}$ completely cancels from the product $M_{\mathbf{q}\lambda}^2 D_\lambda(\mathbf{q}, \omega)$, and the final formula contains no reference to $\Omega_{\mathbf{q}\lambda}$ whatsoever.

The preceding theory was much too simple to find realistic phonon modes in actual metals. The arguments applied only to longitudinal modes of vibration and only at long wavelength. A better description is needed to calculate transverse modes. Dyson's equation for phonons is really a matrix equation, so that the phonon energy denominator is not just a scalar quantity but is derived from the determinant of a matrix.

The reader is referred to Joshi and Rajagopal (1968) for a thorough treatment of phonons in metals. They suggest one method of using the dielectric function for the homogeneous electron gas to calculate realistic phonon modes for metals. If the Fourier transforms of $V_{ii}(\mathbf{r})$ and $V_{ei}(\mathbf{r})$ are $\bar{V}_{ii}(\mathbf{q})$ and $\bar{V}_{ei}(\mathbf{q})$, the screened interaction $\bar{V}_{si}(\mathbf{q})$ potential between ions is given by the effective potential:

$$\bar{V}_{si}(\mathbf{q}) = \bar{V}_{ii}(\mathbf{q}) + \bar{V}_{ei}(\mathbf{q})^2 \frac{P(\mathbf{q})}{\varepsilon(\mathbf{q})} \tag{7.270}$$

The first term is from the direct interaction between ions. The second term is from the interaction mediated by the electrons. The latter term arises when the ions polarize the electrons, and this polarization acts upon other ions. The motivation for choosing this interaction comes from the form of the denominator of the phonon's Green's function, whose poles predict that the phonon frequencies are given by

$$\omega^2 = \Omega_{\mathbf{q}\lambda}^2 + 2\Omega_{\mathbf{q}\lambda}\frac{M_{\mathbf{q}\lambda}^2 P(q)}{\varepsilon(q)} \tag{7.271}$$

The first term on the right, $\Omega_{\mathbf{q}\lambda}^2$, is roughly given by $\phi_{\mu\nu}(q)/M \sim q_\mu q_\nu \bar{V}_{ii}(q)/\rho$. The second term on the right is

$$2\Omega_{\mathbf{q}\lambda}\frac{M_{\mathbf{q}\lambda}^2 P(q)}{\varepsilon(p)} = \frac{q^2}{\rho}\bar{V}_{ei}^2(q)\frac{P(q)}{\varepsilon(q)} \tag{7.272}$$

The two terms combined have the form

$$\omega^2 \sim \frac{q^2}{\rho}\left[\bar{V}_{ii} + \bar{V}_{ei}^2(q)\frac{P(q)}{\varepsilon(q)}\right] \tag{7.273}$$

The term in brackets is just the effective interaction in (7.270). The Green's function derivation has indicated this form to be the appropriate one for the potential between two ions. The effective interaction will be used to solve for the phonon modes in metals. These modes are found by solving a determinantal equation. For real metals it is not a good approximation to express the screened ion potential as $\bar{V}_{ii}(\mathbf{q})/\varepsilon(\mathbf{q})$.

The phonon modes may be obtained in the same fashion as described in (7.249) and (7.254). A dynamical matrix is constructed from the screened ion potential $\phi_{\mu\nu}(\mathbf{q})$ which is used in the determinantal equation (7.254). That procedure was followed by Woll and Kohn (1962) to calculate phonon modes in aluminum. Other calculations are described by Ziman (1960), Sham and Ziman (1963), and Harrison (1966). Other calculations are Na and K by Shukla and Taylor (1976); K, Sn, and Sb by Kay and Reissland (1976); In by Garrett and Swihart (1976); and Na, K, and Rb by Srivastava and Singh (1976).

A significant feature of phonon modes in metals is the *Kohn anomaly* (1959). In the RPA, the electron polarization operator for $\omega = 0$ is ($x = q/2k_F$)

$$P^{(1)}(q) = -\frac{mk_F}{2\pi^2}\left[1 + \frac{1}{2x}(1 - x^2)\ln\left|\frac{1+x}{1-x}\right|\right] \tag{7.274}$$

In the equations for $\omega_\lambda^2(q)$, from $P^{(1)}$ there appears the factor

$$(q^2 - 4k_F^2)\ln(q - 2k_F) \tag{7.275}$$

which is zero for $q = 2k_F$ but has an infinite slope. The phonon modes have a dispersion $d\omega_\lambda(q)/dq$ which is logarithmically divergent at $q = 2k_F$, which is the anomaly.

7.4.2. Electron Self-Energies

The electron–phonon interactions in metals have an important influence upon electron states near the Fermi energy. The energy scale is set by the phonon energies themselves and so is roughly a Debye energy. It is only 20–30 meV in many metals, compared to Fermi degeneracies E_F of several electron volts. The actual self-energy $\Sigma(k, u)$ of electrons from electron–phonon interactions is a small energy, which is negligible compared to E_F. However,

its derivative $d\Sigma(k, u)/du$ is large, so that it makes a large contribution to the electron effective mass. Exactly the opposite behavior was found earlier in Chapter 5 for the electron self-energies from electron–electron interactions: There the energies are large, but the contributions to the effective mass are small.

For the electron–phonon interaction, it is now assumed that the phonon states have been determined satisfactorily, perhaps by the methods already discussed. Having started with a set of bare phonons with frequency $\Omega_{\mathbf{q}\lambda}$ and matrix element $M_{\mathbf{q}\lambda}$ the actual phonon states in the metal have been determined by numerical means. There is available a new set of phonon frequencies $\omega_\lambda(\mathbf{q})$, eigenvectors, and matrix element $\bar{M}_{\mathbf{q}\lambda}$ for electron–phonon interaction, where λ is the mode index: TA, LA, etc. The matrix element is obtained as follows: the screened electron–phonon interaction (7.209) is rewritten in terms of the new matrix elements and frequencies:

$$V_{sp}(\mathbf{q}) = \sum_\lambda \frac{M_\lambda(\mathbf{q})^2}{\varepsilon(q,0)^2} \frac{2\Omega_{\mathbf{q}\lambda}}{\omega^2 - \Omega_{\mathbf{q}\lambda}^2 - 2\Omega_{\mathbf{q}\lambda} P(\mathbf{q}) M_\lambda^2/\varepsilon(\mathbf{q})} \tag{7.276}$$

$$= \sum_\lambda \bar{M}_\lambda^2(\mathbf{q}) \frac{2\omega_\lambda(\mathbf{q})}{\omega^2 - \omega_\lambda^2(\mathbf{q})} = \sum_\lambda \bar{M}_\lambda^2(\mathbf{q}) \bar{D}^{(0)}(\mathbf{q}, \omega) \tag{7.277}$$

The frequency dependence in $P(\mathbf{q}, \omega)$ and $\varepsilon(\mathbf{q}, \omega)$ is ignored since it has no influence at small frequencies. The preceding equation serves to define the renormalized matrix element from (7.258),

$$\bar{M}_\lambda(\mathbf{q}) = \frac{M_\lambda(\mathbf{q})}{\varepsilon(\mathbf{q})} \left(\frac{\Omega_{\mathbf{q}\lambda}}{\omega_\lambda(\mathbf{q})}\right)^{1/2} = \left(\frac{\hbar}{2\rho\omega_\lambda(\mathbf{q})}\right)^{1/2} \mathbf{q} \cdot \hat{\xi}_{\mathbf{q}\lambda} \frac{\bar{V}_{ei}(\mathbf{q})}{\varepsilon(\mathbf{q})} \tag{7.278}$$

which is expressed in terms of the renormalized frequencies $\omega_\lambda(\mathbf{q})$. This renormalized version of the screened electron–phonon interaction is exactly what we would obtain from a Hamiltonian of the form

$$H = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C_{\mathbf{p}\sigma}^\dagger C_{\mathbf{p}\sigma} + \sum_{\mathbf{q}\lambda} \left[\omega_\lambda(\mathbf{q}) a_{\mathbf{q}\lambda}^\dagger a_{\mathbf{q}\lambda} + \frac{\bar{M}_\lambda(\mathbf{q})}{\sqrt{\nu}} A_{\mathbf{q}\lambda} \sum_{\mathbf{p}\sigma} C_{\mathbf{p}+\mathbf{q},\sigma}^\dagger C_{\mathbf{p}\sigma} \right] \tag{7.279}$$

Equation (7.279) is just a simple electron–phonon Hamiltonian using the renormalized phonon frequencies and matrix elements. Its significant feature is that electron–electron interactions are omitted. When they are included, they serve to screen the electron–phonon interactions. This phenomenon was used to derive the renormalized frequencies and matrix elements. Adding electron–electron interactions again would be double counting. It would give a second renormalization which is incorrect. Use the effective Hamiltonian (7.279), with renormalized frequencies, and do not include screening again.

The self-energy of the electron, from electron–phonon interactions, is calculated using the renormalized phonons and matrix elements. The only diagram which is ever calculated is the one-phonon bubble diagram:

$$\Sigma(\mathbf{k}, ik_n) = -\frac{1}{\beta\nu} \sum_{\mathbf{q}\lambda, i\omega_m} \bar{M}_\lambda(\mathbf{q})^2 \mathscr{G}^{(0)}(\mathbf{k}+\mathbf{q}, ik_n + i\omega_m) \bar{D}_\lambda(\mathbf{q}, i\omega_m)$$

$$= \sum_\lambda \int \frac{d^3q}{(2\pi)^3} \bar{M}_\lambda(\mathbf{q})^2 \frac{1}{\beta} \sum_{i\omega_m} \frac{2\omega_\lambda(\mathbf{q})}{\omega_m^2 + \omega_\lambda(\mathbf{q})^2} \frac{1}{ik_n + i\omega_m - \xi_{\mathbf{k}+\mathbf{q}}} \tag{7.280}$$

The one-phonon self-energy makes only a small contribution to the energy of the electron. Higher-order terms, involving two or more phonons, are neglected on the basis that their contribution is even smaller, as first shown by Migdal (1958). The one-phonon self-energy diagram is evaluated by making several approximations. The free-particle propagator $\mathscr{G}^{(0)}$ is used for the electron Green's function. The electron–phonon matrix elements $\bar{M}_\lambda(\mathbf{q})$ are treated as functions only of $\mathbf{q}$, not of ω. The frequency dependence is ignored in the longitudinal dielectric function $\varepsilon(\mathbf{q}, i\omega)$. With these approximations, the Matsubara summation is the same one which has been evaluated previously in Sec. 3.5 for unperturbed electron and phonon Green's functions:

$$\Sigma(\mathbf{k}, ik_n) = \sum_\lambda \int \frac{d^3q}{(2\pi)^3} \bar{M}_\lambda(\mathbf{q})^2 \left[\frac{n_B(\omega_\lambda) + 1 - n_F(\xi_{\mathbf{k}+\mathbf{q}})}{ik_n - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})} + \frac{n_B(\omega_\lambda) + n_F(\xi_{\mathbf{k}+\mathbf{q}})}{ik_n - \xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})} \right] \tag{7.281}$$

The retarded function is obtained by $ik_n \to u + i\delta$. The discussion is confined to zero-temperature properties, so set $n_B = 0$. Consider the expression

$$\Sigma(\mathbf{k}, u) = \sum_\lambda \int \frac{d^3q}{(2\pi)^3} \bar{M}_\lambda(\mathbf{q})^2 \left[\frac{1 - n_F(\xi_{\mathbf{k}+\mathbf{q}})}{u - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q}) + i\delta} + \frac{n_F(\xi_{\mathbf{k}+\mathbf{q}})}{u - \xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q}) + i\delta} \right] \tag{7.282}$$

The self-energy is customarily evaluated using the Brillouin–Wigner form of perturbation theory. The energy u is retained as a separate parameter, and the self-energy is calculated as a function of both u and k.

The variation of $\Sigma(k, u)$ with respect to u is far more important than with respect to k. The variation with respect to u is on the scale of a phonon energy, while the variation with respect to k is on the scale of k_F. An order of magnitude estimate of these derivatives is

$$\frac{\partial \Sigma}{\partial u} \sim \frac{\Sigma}{\omega_D}, \qquad \frac{\partial \Sigma}{\partial \xi_k} \sim \frac{\Sigma}{E_F} \tag{7.283}$$

where ω_D is the Debye energy of the solid. The derivative of $\Sigma(k, u)$ with respect to u is larger than the derivative with respect to ξ_k by the factor $E_F/\omega_D \sim 10^2$. It is usually possible to neglect $\partial\Sigma/\partial\xi_k$, so that the effective mass and renormalization factors are given by the same expression:

$$\left(\frac{m}{m^*}\right) = Z_F = \left(1 - \frac{\partial \Sigma}{\partial u}\right)^{-1}_{u=0, k=k_F} \tag{7.284}$$

The electron self-energy $\Sigma(k, u)$ from electron–phonon interactions has appreciable value only for u within a Debye energy of the Fermi surface, or $-\omega_D < u < \omega_D$. In this narrow energy range, k hardly changes from k_F. For metals with a spherical Fermi surface, the k dependence of $\Sigma(k, u)$ is unimportant, and the k dependence is often even suppressed in the notation: $\Sigma(u) \equiv \Sigma(k_F, u)$. When the metal has its Fermi sphere cut by Bragg planes, so that the Fermi surface is divided into pockets, the self-energy may depend on the position on the Fermi surface in the region near these Bragg planes. In that case one should retain the $\mathbf{k}$ dependence.

The two terms in the integrand of (7.282) which contain the factor n_F are far more important than the factor which contains the 1. It is natural to separate Σ into two terms:

$$\Sigma = \Sigma^{(a)} + \Sigma^{(b)} \tag{7.285}$$

$$\Sigma^{(a)}(\mathbf{k}, u) = \sum_\lambda \int \frac{d^3q}{(2\pi)^3} n_F(\xi_{\mathbf{k}+\mathbf{q}}) \bar{M}_\lambda(\mathbf{q})^2 \times \left[\frac{1}{u - \xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q}) + i\delta} - \frac{1}{u - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q}) + i\delta} \right] \tag{7.286}$$

$$\Sigma^{(b)}(\mathbf{k}, u) = \sum_\lambda \int \frac{d^3q}{(2\pi)^3 \bar{M}_\lambda(\mathbf{q})^2} \frac{1}{u - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q}) + i\delta} \tag{7.287}$$

The term $\Sigma^{(b)}$ is a rather dull function of its two arguments. Not only is this self-energy term small in magnitude but so are its derivatives with respect to $\mathbf{k}$ and u. It makes a negligible contribution to the energy, effective mass, or other properties of the electron. An estimate of this term can be obtained in the following way. Since the chemical potential does not enter the evaluation of this self-energy, it behaves as the self-energies calculated for one particle in an empty band. It contributes to the self-energy for states near to the bottom of the occupied band, but is quite small for electrons near to the Fermi energy. Only the latter electrons matter, so $\Sigma^{(b)}$ can be neglected.

The other self-energy term, $\Sigma^{(a)}$ in (7.286), is also rather small in value. Its importance comes from its contribution to the electron effective mass. The self-energy has a strong dependence upon energy for $u \sim 0$, which is near the Fermi energy. Consider the real part of $\partial\Sigma^{(a)}/\partial u$

$$\mathrm{Re}\left(\frac{\partial \Sigma^{(a)}}{\partial u}\right)_{u=0} = -\sum_\lambda \int \frac{d^3q}{(2\pi)^3} n_F(\xi_{\mathbf{k}+\mathbf{q}}) \bar{M}_\lambda(\mathbf{q})^2 \left[\frac{1}{[\xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})]^2} - \frac{1}{[\xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})]^2} \right] \tag{7.288}$$

The integral is evaluated by integrating by parts on the angular variable $\nu = \cos\theta = \hat{p} \cdot \hat{q}$, where the two factors are

$$\int u dv = uv - \int v du \tag{7.289}$$

$$v = -\frac{m}{kq}(\xi_{\mathbf{k}+\mathbf{q}} \pm \omega_\lambda)^{-1}, \qquad dv = d\nu(\xi_{\mathbf{k}+\mathbf{q}} \pm \omega_\lambda)^{-2}$$

$$u = n_F(\xi_{\mathbf{k}+\mathbf{q}}), \qquad du = \frac{kq}{m} d\nu \left[\frac{d}{d\xi} n_F(\xi_{\mathbf{k}+\mathbf{q}}) \right]$$

The integration by parts gives, with $d^3q = 2\pi q^2 dq d\nu$

$$\mathrm{Re}\left(\frac{\partial \Sigma^{(a)}}{\partial u}\right)_{u=0} = \sum_\lambda \int \frac{q^2 dq}{(2\pi)^2} \bar{M}_\lambda(\mathbf{q})^2 \left[n_F(\xi_{\mathbf{k}+\mathbf{q}}) \left(\frac{1}{\xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})} - \frac{1}{\xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})} \right)_{\nu=-1}^{\nu=1} - \int_{-1}^{1} d\nu \left(\frac{d}{d\xi} n_F(\xi_{\mathbf{k}+\mathbf{q}}) \right) \left(\frac{1}{\xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})} - \frac{1}{\xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})} \right) \right] \tag{7.290}$$

The first term in (7.290) is smaller than the second by a factor ω_D/E_F and may be neglected. The second term is the important one. The factor $dn_F(\xi)/d\xi = -\delta(\xi)$ at zero temperature. Changing the integration variable to $\mathbf{p} = \mathbf{k} + \mathbf{q}$ gives, for this last term alone,

$$\mathrm{Re}\left(\frac{\partial\Sigma^{(a)}}{\partial u}\right)_{u=0} = -2\sum_{\lambda}\int\frac{d^3p}{(2\pi)^3}\delta(\xi_{\mathbf{p}})\frac{\bar{M}_\lambda(\mathbf{k}-\mathbf{p})^2}{\omega_\lambda(\mathbf{k}-\mathbf{p})} \tag{7.291}$$

$$= -\frac{2}{(2\pi)^3}\sum_{\lambda}\int\frac{d^2p}{v_F}\frac{\bar{M}_\lambda(\mathbf{k}-\mathbf{p})^2}{\omega_\lambda(\mathbf{k}-\mathbf{p})} \tag{7.292}$$

where $d^3p = 2\pi p^2 dp d\nu$ and $\int d\nu\delta(\xi) = 1/(pv_F)$, which leaves $d^2p = 2\pi p dp$. The remaining integration is over the two-dimensional area of the Fermi surface. For spherical Fermi surfaces, this factor is $d^2p/v_F = mk_F d\Omega_p$, where Ω_p is solid angle, but most Fermi surfaces are not spherical. The term $[\partial\,\mathrm{Re}\,\Sigma^{(a)}/du]$ is quite large because of the factor ω_λ^{-1}, since phonon energies are small. This important expression was first derived by Nakajima and Watabe (1963). It is called λ,

$$\lambda(\hat{k}) = -\mathrm{Re}\left(\frac{\partial\Sigma^{(a)}}{\partial u}\right)_{u=0} = \frac{2}{(2\pi)^3}\sum_{\lambda}\int\frac{d^2p}{v_F}\frac{\bar{M}_\lambda(\mathbf{k}-\mathbf{p})^2}{\omega_\lambda(\mathbf{k}-\mathbf{p})} \tag{7.293}$$

which is somewhat unfortunate, since λ is also used to designate the summation over the polarization modes of the phonon system. No confusion should arise over these quite different parameters, however.

The parameter $\lambda(\hat{k})$, as defined in (7.293), is a function of the position on the Fermi surface of the metal. The point $\mathbf{k}$ is the reference point, and the average is taken over other points $\mathbf{p}$ on the Fermi surface. Quite often the average value of λ is given for a metal, which is obtained by averaging $\lambda(\hat{k})$ over the Fermi surface:

$$\lambda = \frac{\int d^2k\lambda(\hat{k})}{\int d^2k} \tag{7.294}$$

The other contribution to the effective mass of a particle at the Fermi surface is from the derivative of $\Sigma(\mathbf{k}, u)$ with respect to ξ_k. However, $\partial\Sigma/\partial\xi_k$ is much smaller than λ by a factor ω_D/E_F, so that the most significant term is λ. The other terms are usually entirely neglected, and the effective mass of electrons at the Fermi energy, from electron–phonon contributions, is given by

$$\left(\frac{m^*}{m}\right) = 1 + \lambda \tag{7.295}$$

The first calculations of this quantity were by Swihart *et al.* (1965), and by Ashcroft and Wilkins (1965), and many have been done since. The quantity λ can be determined from electron tunneling experiments, so that experimental values are available. Both theory and experiment have been reviewed by Grimvall (1981). His suggested values are shown in Table 7.2 for some nontransition metals. Some of the values are quite large—they are 1.6 and 1.5 for mercury and lead. These large values show that the electron–phonon coupling is large for these metals. This observation is significant for the theory of superconductivity, which in elemental metals is due to electron–phonon coupling as described in Chapter 10. Metals which have a low value of λ, such as the alkali metals, are not superconducting.

Table 7.2 Electron phonon mass enhancement λ.

Li	Be		
0.41 ± 0.15	0.24 ± 0.05		
Na	Mg	Al	
0.16 ± 0.04	0.35 ± 0.04	0.43 ± 0.05	
K	Zn	Ga	
0.13 ± 0.03	0.37 ± 0.05	0.97 ± 0.05	
Rb	Cd	In	Sn
0.16 ± 0.04	0.40 ± 0.05	0.8	0.72
Cs	Hg	Tl	Pb
0.15 ± 0.04	1.6	0.8	1.5

From Grimvall (1981).

The phonon density of states is defined to be

$$F(\omega) = \sum_{\lambda} \int \frac{d^3q}{(2\pi)^3} \delta[\omega - \omega_\lambda(\mathbf{q})] \tag{7.296}$$

where the summation λ is over different modes, and the integration is over the Brillouin zone. McMillan (1968) has introduced another function, which is called $\alpha^2 F(\omega)$. It is treated as a single function, although the notation implies incorrectly that it is the product of two functions. It is defined as

$$\alpha_{\mathbf{k}}^2 F(\omega) = \frac{1}{(2\pi)^3} \int \frac{d^2p}{v_F} \bar{M}_\lambda(\mathbf{k} - \mathbf{p})^2 \delta[\omega - \omega_\lambda(\mathbf{k} - \mathbf{p})] \tag{7.297}$$

$$\lambda(\hat{k}) = 2 \int_0^{\omega_D} \frac{d\omega}{\omega} \alpha_{\mathbf{k}}^2 F(\omega) \tag{7.298}$$

$\alpha_{\mathbf{k}}^2 F(\omega)$ is the frequency spectrum one obtains by starting at a point on the Fermi surface $\mathbf{k}$ and integrating over all other points on the Fermi surface $\mathbf{p}$. It will vary from point to point on the Fermi surface of a metal. The same quantity without the subscript $\mathbf{k}$ is the average over all different points on the Fermi surface:

$$\alpha^2 F(\omega) = \int \frac{d^2k}{v_F} \alpha_{\mathbf{k}}^2 F(\omega) \Big/ \int \frac{d^2k}{v_F} \tag{7.299}$$

Nowadays it is calculated with a computer, using realistic electron wavefunctions, realistic phonon eigenstates and frequencies, and realistic matrix elements. The quantity $\alpha^2 F(\omega)$ can be obtained experimentally, as will be explained in Chapter 10, from electron tunneling experiments. As an example, $\alpha^2 F(\omega)$ for Pb and Sn are given in Fig. 7.13. These results are from the experimental data of Rowell *et al.* (1969a, 1971). The dashed line in Fig. 7.13(a) is $F(\omega)$ for Pb, which is similar to $\alpha^2 F(\omega)$ because the matrix elements vary only by about a factor of 2 over the various phonon states. Other physical quantities may be calculated from $\alpha^2 F(\omega)$. For example, the lifetime of electrons on the Fermi surface is

$$\frac{1}{\tau(\mathbf{k})} = 4\pi \int_0^{\omega_D} d\omega \alpha_{\mathbf{k}}^2 F(\omega) \left(\frac{1}{e^{\beta\omega} + 1} + \frac{1}{e^{\beta\omega} - 1} \right) \tag{7.300}$$

Note that $1/\tau(\mathbf{k})$ is zero at zero temperature but is nonzero for a nonzero temperature. This lifetime is between scatterings of the electron by phonon emission or absorption. It is not the same lifetime which enters a calculation of the resistivity, for example. These differences are

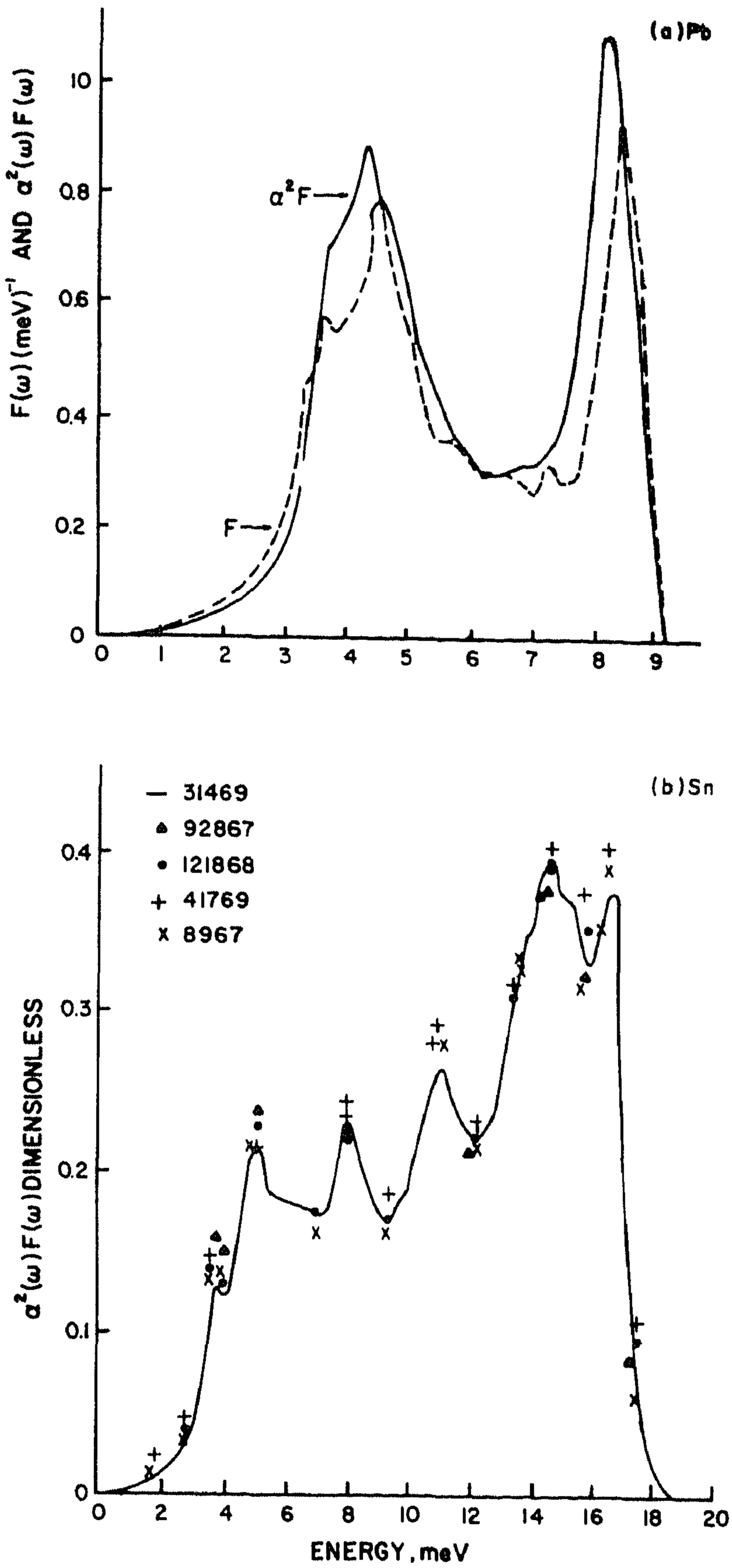


FIGURE 7.13 The energy dependence of the electron–phonon coupling $\alpha^2F(\omega)$ in (a) lead and (b) tin as determined by electron tunneling in superconductors. In (a) the dashed line is the phonon density of state $F(\omega)$ alone. In (b) the points indicate the spread of values for different tunnel junctions. *Source*: Rowell *et al.* (1969a, 1971) (used with permission).

treated in Chapter 8. The electron lifetime in (7.300) is derived from the imaginary part of the retarded self-energy in (7.281):

$$\frac{1}{\tau(\mathbf{k})} = -2\,\mathrm{Im}[\Sigma(\mathbf{k}, u=0)] = 2\pi \sum_\lambda \int \frac{d^3q}{(2\pi)^3} \bar{M}_\lambda(\mathbf{q})^2$$
$$\times \{[1 + n_B(\omega_\lambda) - n_F(\xi_{\mathbf{k}+\mathbf{q}})]\delta(\xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q}))$$
$$+ [n_B(\omega_\lambda) + n_F(\xi_{\mathbf{k}+\mathbf{q}})]\delta(\xi_{\mathbf{k}+\mathbf{q}})]\delta(\xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q}))\} \tag{7.301}$$

Change variables to $\mathbf{p} = \mathbf{k} + \mathbf{q}$, and the integral is

$$\frac{1}{\tau(\mathbf{k})} = 2\pi \sum_\lambda \int \frac{d^3p}{(2\pi)^3} \bar{M}_\lambda(\mathbf{k}-\mathbf{p})^2 \{n_B(\omega_\lambda(\mathbf{k}+\mathbf{p})) + n_F(\omega_\lambda(\mathbf{k}-\mathbf{p}))\}$$
$$\times \{\delta(\xi_\mathbf{p} - \omega_\lambda(\mathbf{k}-\mathbf{p})) + \delta(\xi_\mathbf{p} - \omega_\lambda(\mathbf{k}-\mathbf{p}))\} \tag{7.302}$$

The phonon energy can be neglected in the delta functions $\delta(\xi_\mathbf{p} \pm \omega_\lambda(\mathbf{k}-\mathbf{p}))$ since their energy is small compared the the Fermi energy. Then the two delta functions are alike, and each just scatter the electron to another point in the Fermi surface

$$\frac{1}{\tau(\mathbf{k})} = \frac{4\pi}{(2\pi)^3} \sum_\lambda \int \frac{d^2p}{v_F} \bar{M}_\lambda(\mathbf{k}-\mathbf{p})^2 \{n_B(\omega_\lambda) + n_F(\omega_\lambda)\} \tag{7.303}$$

which is the same as (7.300). Results for aluminum and lead have been reported by Tomlinson and Carbotte (1976). The same calculation for nonzero u gives

$$\frac{1}{\tau(\mathbf{k}, u)} = -2\,\mathrm{Im}[\Sigma(\mathbf{k}, u)] \tag{7.304}$$

$$= 2\pi \int_0^{\omega_D} d\omega \alpha_\mathbf{k}^2 F(\omega)[2n_B(\omega) + n_F(\omega + u) + n_F(\omega - u)] \tag{7.305}$$

Equations (7.300) and (7.305) are the formulas for calculating $\mathrm{Im}(\Sigma)$.

Next a formula is needed for $\mathrm{Re}[\Sigma(k, u)]$. From (7.286), integrate by parts on the $\nu = \cos\theta$ variable in the same fashion as was done in deriving (7.290):

$$\int_{-1}^{1} d\nu n_F(\xi_{\mathbf{k}+\mathbf{q}}) \left[\frac{1}{u - \xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})} - \frac{1}{u - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})}\right]$$
$$= -\frac{m}{kq} n_F(\xi_{\mathbf{k}+\mathbf{q}}) \ln\left|\frac{u - \xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})}{u - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})}\right|_{\nu=-1}^{\nu=1}$$
$$+ \int_{-1}^{1} d\nu \frac{dn_F(\xi)}{d\xi} \ln\left|\frac{u - \xi_{\mathbf{k}+\mathbf{q}} + \omega_\lambda(\mathbf{q})}{u - \xi_{\mathbf{k}+\mathbf{q}} - \omega_\lambda(\mathbf{q})}\right|$$
$$\approx -\ln\left|\frac{u + \omega_\lambda(\mathbf{q})}{u - \omega_\lambda(\mathbf{q})}\right| \int_{-1}^{1} d\nu \delta(\xi_{\mathbf{k}+\mathbf{q}}) \tag{7.306}$$

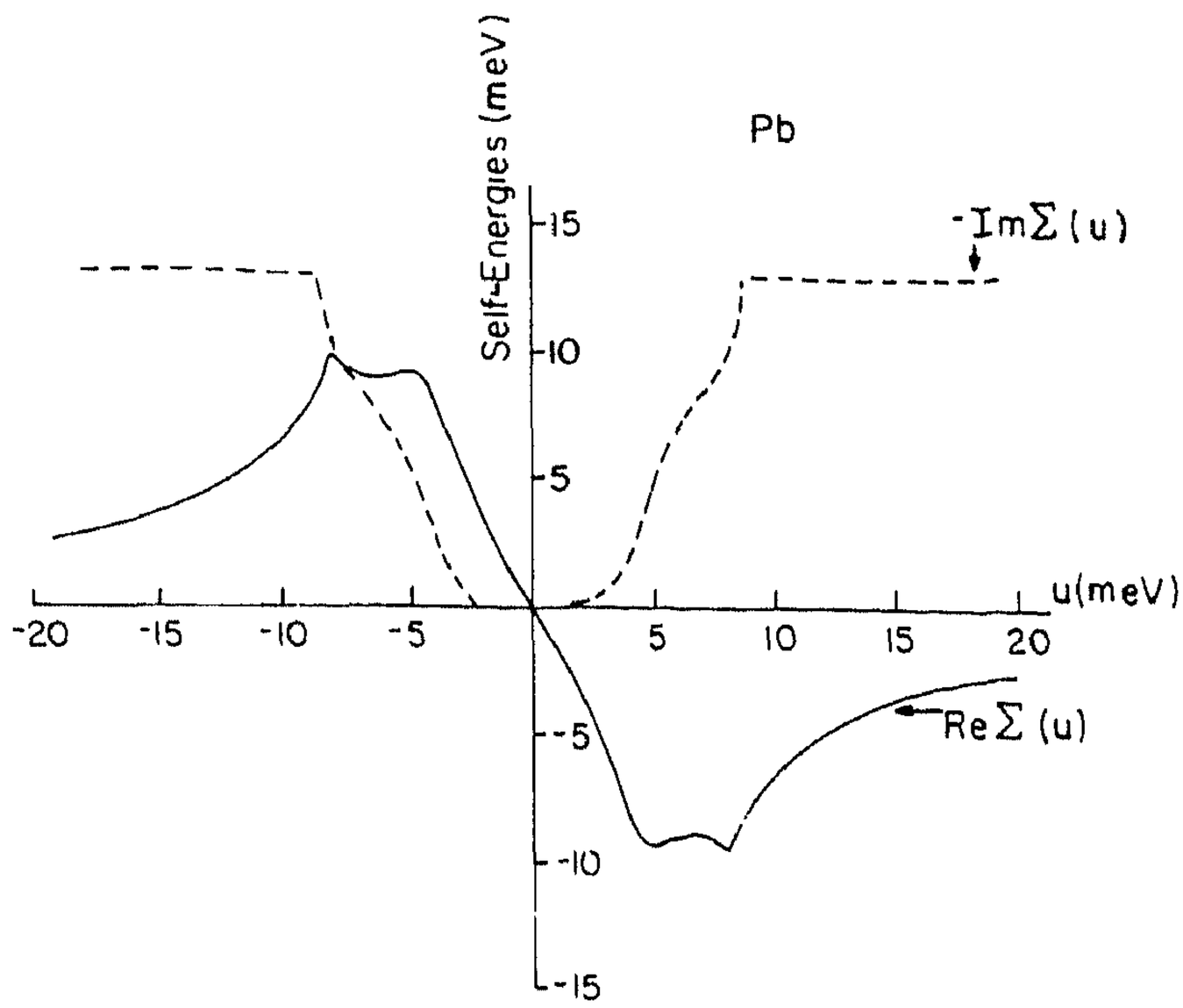


FIGURE 7.14 The real and imaginary self-energies at zero temperature of an electron in lead, from the electron–phonon interaction. *Source*: Grimvall (1981) (used with permission).

The term $dn_F(\xi)/d\xi = -\delta(\xi)$ is the important one, and the other may be neglected in comparison. A careful analysis shows it to be smaller by the factor ω_D/E_F. The result for zero temperature is:

$$\begin{aligned}\mathrm{Re}[\Sigma(\mathbf{k}, u)] &= -\sum_\lambda \int \frac{d^3p}{(2\pi)^3}\delta(\xi_\mathbf{p})\bar{M}_\lambda(\mathbf{k}-\mathbf{p})^2 \ln\left|\frac{u+\omega_\lambda(\mathbf{k}-\mathbf{p})}{u-\omega_\lambda(\mathbf{k}-\mathbf{p})}\right| \\ &= -\int_0^{\omega_D} d\omega\alpha_\mathbf{k}^2(\omega)\ln\left|\frac{\omega+u}{\omega-u}\right| \qquad (7.307)\end{aligned}$$

This expression vanishes at $u = 0$ and is positive for $u < 0$ and negative for $u > 0$. The usual procedure is to find $\alpha^2F(\omega)$ from superconducting tunneling experiments. It is a property of both the normal metal and the superconductor. It is used in (7.305) and (7.307) to find the electron self-energy from electron–phonon interactions. Corresponding experimental information may be obtained from electron tunneling in normal metals, as shown by Rowell *et al.* (1969b). Their results for Pb are shown in Fig. 7.14. The self-energy has no **k** dependence since they use an averaged value of $\alpha^2F(\omega)$. $\mathrm{Re}[\Sigma(u)]$ is quite small—of the order of milli-electron volts. However, it has a very steep slope at point $u = 0$, which gives $d\Sigma/du$ its large value.

PROBLEMS

1. The polaron self-energy $-\alpha\omega_0$ contains some kinetic energy of the electron and some potential energy. Use the Rayleigh–Schrödinger wave function to show these have the ratio of 1 : 3 at $p = 0$.

2. Show that the total momentum of the polaron in (7.18) commutes with the Hamiltonian in (7.2).

3. Calculate m/m^* for the Fröhlich polaron in the Tamm–Dancoff approximation.

4. Use the two-phonon wave function, in Rayleigh–Schrödinger perturbation theory, to find the α^2 term for the number of phonons in the polaron cloud at $p = 0$. Be sure your wave function is normalized to unity. You should get $-0.1036\alpha^2$, so correlations reduce the number of phonons in the cloud.

5. Calculate $\Sigma_{RS}(p)$ at nonzero temperature for all values of ε_p. Also calculate its imaginary part, which is the imaginary part of the retarded self-energy with $E = \varepsilon_p$, which is also half the mean-free-path for one-phonon emissions or absorptions.

6. Use the answer of Problem 5 to plot, on a piece of graph paper, $\Sigma_{RS}(\varepsilon_p)$ for $\alpha = 1$ and $0 < \varepsilon_p < \omega_0$. Evaluate the self-energy in the Tamm–Dancoff approximation for the same parameters, and plot it on the same set of graph paper.

7. Consider the one-phonon self-energy in Rayleigh–Schrödinger perturbation theory at nonzero temperatures:

$$\Sigma_{RS}(p) = \int \frac{d^3q}{(2\pi)^3} M_q^2 \left(\frac{N_q + 1}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p+q}} - \omega_q} + \frac{N_q}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p+q}} + \omega_q} \right) \tag{7.308}$$

At high temperatures, make the approximation $N_q \approx k_B T/\omega_q$, and keep only the term proportional to $k_B T$. Evaluate it by assuming a Debye spectrum and a piezoelectric matrix element (1.256). You should find that the self-energy vanishes when v_p is less than the speed of sound.

8. Minimize the ground state energy, in strong coupling theory, with the variational wave function $\phi(r) = A\exp(-\beta r)$. Show that the energy is $-0.0980\alpha^2\omega_0$, which is not as low as found with a Gaussian wave function.

9. In strong coupling theory, calculate the energy of the lowest p state by using the variational wave function $\phi(r) = Ar\cos(\theta)\exp(-\beta^2 r^2)$, where β is a variational parameter.

10. In the limit of small wave vectors, the interaction between an electron and the homogeneous electron gas has the form

$$v_q\left(\frac{1}{\varepsilon} - 1\right) \to \frac{v_q \omega_p^2}{\omega^2 - \omega_p^2} \tag{7.309}$$

The electron–plasmon interaction is identical in form to the Fröhlich polaron Hamiltonian. What is the equivalent α for the electron gas? Find the numerical values for metallic sodium ($r_s = 3.96$) and aluminum ($r_s = 2.07$).

11. Consider a small-polaron Hamiltonian which has the phonons localized on each site. This model could apply to molecular crystals, where the molecular vibrations are highly localized:

$$H = \sum_i \left[J \sum_\delta C_{i+\delta}^\dagger C_i + \omega_0 a_i^\dagger a_i + \lambda\omega_0 C_i^\dagger C_i (a_i + a_i^\dagger) \right] \tag{7.310}$$

Show that this may be derived from (7.95) with the choice $M_q =$ constant. Find the canonical transformation which diagonalizes the last two terms in the Hamiltonian. Use the transformed Hamiltonian to calculate:

(a) The thermally averaged amplitude and probability that a hop occurs without any change in the number of phonons at each site.
(b) The thermally averaged amplitude and probability that the hop occurs without any change in the number of phonons but that they may be exchanged between the two sites.

12. Show that the effective interaction (7.173) with the dielectric function (7.174) may be written as

$$V_{\text{eff}} = v_q\left[\frac{1}{\varepsilon_\infty - v_q P} - \frac{\Omega(q)^2}{\Omega(q)^2 - \omega^2}\left(\frac{1}{\varepsilon_\infty - v_q P} - \frac{1}{\varepsilon_0 - v_q P}\right)\right]$$

where $\Omega(q)$ is the renormalized phonon modes of (7.220). At $\omega = 0$, then V_{eff} is $v_q/(\varepsilon_0 - v_q P)$, and at $\omega \to \infty$ it is $v_q/(\varepsilon_\infty - v_q P)$ (Appel, 1966).

13. Prove that the following two expressions are identical for polar scattering by optical phonons

$$V_{\text{eff}} = \frac{v_q}{\varepsilon_{\text{total}}} = \frac{W}{1 - WP} \tag{7.311}$$

where $\varepsilon_{\text{total}}$ is given in (7.173) and W is given in (7.194).

14. The Raman scattering of Fig. 7.11 occurs because the light couples nonlinearly to the phonons but not to the plasmons. Assume a term in the Hamiltonian of the form

$$I\sum_{\mathbf{k}\mathbf{k}'}[b_{\mathbf{k}}^\dagger b_{\mathbf{k}'} A_{\mathbf{k}'-\mathbf{k}} + \text{hc}] \tag{7.312}$$

which describes the scattering of light from $\mathbf{k}' \to \mathbf{k}$ with photon operators $b_{\mathbf{k}}^\dagger b_{\mathbf{k}'}$ and phonon operators $A_{\mathbf{q}} = a_{\mathbf{q}} + a_{-\mathbf{q}}^\dagger$.

(a) Derive a correlation function which describes the Raman scattering to order I^2.
(b) Solve it at long wavelengths to get the Raman intensity of the modes $\omega_\pm$ in (7.214) as a function of n_0.

15. Compare Rayleigh–Schrödinger and Brillouin–Wigner perturbation theory by plotting both theories using the self-energy in (7.233):

$$\text{RS:} \qquad E(p) = \xi_p + \text{Re}[\Sigma(\xi_p)] \tag{7.313}$$

$$\text{BW:} \qquad E(p) = \xi_p + \text{Re}\{\Sigma[E(p)]\} \tag{7.314}$$

Use the parameters $f_0 = 10$ meV and $\omega_0 = 30$ meV. Which theory appears more reasonable?

16. Use the RPA for $\varepsilon_2(q, \omega)$ to estimate the damping of phonons from electron–hole creation in a metal.

17. Prove that the sum rule $\int d\omega\omega\alpha^2 F(\omega)$ is independent of phonon properties (McMillan, 1968). Evaluate the sum rule analytically for metallic hydrogen, and estimate the formula numerically.

18. Derive Eq. (7.305) for $\tau(k, u)^{-1}$. Show that it goes to a constant at $u \to \infty$.

19. Use (7.305) and (7.307) to show at zero temperature that $\text{Re}[\Sigma(u)]$ and $\text{Im}[\Sigma(u)]$ obey a Kramers–Kronig relation. Calculate by direct integration of (5.297) and (5.298).

20. Show that in the jellium model of a metal, with the ions as point charges, the longitudinal modes at long wavelength are described by a total dielectric function [see (7.174)]:

$$\varepsilon_{\text{total}} = \varepsilon_1\left(1 - \frac{\omega_{ip}^2}{\omega^2}\right) - v_q P(q, \omega) \tag{7.315}$$

With this choice, one can write the effective interaction in (7.173) as the correct electron–electron and electron–phonon parts.

21. If $\Sigma(T, k, \varepsilon)$ is the retarded self-energy at temperature T, show that the electron self-energy in a metal from electron–phonon interactions obeys

$$\mathrm{Re}[\Sigma(T, k, \varepsilon)] = \int_{-\infty}^{\infty} d\varepsilon' \left[-\frac{dn_F(\varepsilon')}{d\varepsilon'} \right] \mathrm{Re}[\Sigma(0, k, \varepsilon - \varepsilon')] \tag{7.316}$$

22. Use (7.300) to show that at high temperatures $1/\tau = 2\pi\lambda k_B T$.

Chapter 8

dc Conductivities

8.1. ELECTRON SCATTERING BY IMPURITIES

This chapter is mostly concerned with the methods for calculating the electrical conductivity. Four different methods are discussed: (1) solving the Boltzmann equation, (2) evaluating the Kubo formula for the current–current correlation function, (3) evaluating the force–force correlation function, and (4) solving the quantum Boltzmann equation. For scattering from fixed impurities they all give the same answer. For scattering by phonons two different answers are obtained. One is called the Ziman (1960) formula, and the other the Holstein (1964) formula. Two criteria are important in comparing these methods: which is the easiest to use, and which gives the most accurate answer?

The electrical resistivity, or conductivity, from impurity scattering is an important topic. From the experimental viewpoint, all solids have impurities which make a contribution to the total resistivity. In many metals and semiconductors, the low-temperature resistivity is dominated by impurities, since all other contributions are temperature dependent and vanish at low temperature. In metals, the resistivity from impurity scattering is largely temperature independent, except for temperature variations on the scale of the Fermi temperature $T_F = E_F/k_B$. The subject is important, on the theoretical side, because it was one of the earliest evaluations of the Kubo formula. The importance of vertex corrections became apparent. Indeed, the derivation showed that vertex corrections are usually very important and should be assumed important until shown otherwise. This conclusion, and message, continues to be relevant even for calculations of other quantities.

The present chapter is really about vertex corrections.

8.1.1. Boltzmann Equation

The electrical resistivity from impurity scattering is easily derived by using the Boltzmann equation. This derivation is presented for several reasons. First, it is the easiest way to get the answer. Second, the resistivity was first found this way; the Green's function evaluation of the Kubo formula only confirmed the result known earlier from transport theory. The Green's function derivation is complicated and subtle, and it is useful to know and believe the right answer in order to recognize it when it is finally derived.

Our objective is to derive a formula for the electrical resistivity with the least possible fuss. The simplest possible model is adopted for the solid. It is a homogeneous system except for randomly located impurities. The electron states are plane waves except for occasional scattering from isolated impurities. The impurities are very dilute, so that interference between successive scatterings can be neglected. In the Boltzmann theory, the electrons are described by a classical distribution function $f(\mathbf{r}, \mathbf{k}, t)$. The time rate of change of this distribution function is governed by the Boltmann equation

$$0 = \frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \frac{\partial \mathbf{k}}{\partial t} \cdot \nabla_k f + \left(\frac{df}{dt}\right)_{\text{collisions}} \tag{8.1}$$

The last term is the time rate of change due to collisions with the impurities. There is no $\mathbf{r}$ dependence in $f(\mathbf{r}, \mathbf{k}, t)$ since the material is assumed to be homogeneous. Also, for the dc conductivity, there is no time dependence. The system has a weak external electric field and the current flows in a steady-state fashion. The distribution function is only a function of wave vector $f(\mathbf{k})$ and obeys the equation

$$0 = \frac{\partial \mathbf{k}}{\partial t} \cdot \nabla_k f + \left(\frac{df}{dt}\right)_{\text{collisions}} \tag{8.2}$$

In a solid, the factor $\partial \mathbf{k}/\partial t$ is equivalent to an acceleration which is equal to the forces on the electron (see Kittel, 1963):

$$\frac{\partial \mathbf{k}}{\partial t} = -e\mathbf{E} - \frac{e}{c}\mathbf{v} \times \mathbf{H}_0 \tag{8.3}$$

In the present problem, there is an electric field $\mathbf{E}$ and no magnetic field $\mathbf{H}_0 = 0$, so that

$$e\mathbf{E} \cdot \nabla_k f(\mathbf{k}) = \left(\frac{df}{dt}\right)_{\text{collisions}} \tag{8.4}$$

The collision term is the most interesting. It is evaluated in the *relaxation time approximation.* This approximation assumes that collisions seek to return the system to the equilibrium configuration $f_0(\mathbf{k})$, which is the configuration the system would have in the absence of an electric field. The rate of change of $f(\mathbf{k})$ due to collisions is assumed to be proportional to the degree that $f(\mathbf{k})$ is different from $f_0(\mathbf{k})$:

$$\left(\frac{df}{dt}\right)_{\text{collisions}} = -\frac{[f(\mathbf{k}) - f_0(\mathbf{k})]}{\tau_t(\mathbf{k})} \tag{8.5}$$

$$f_0(\mathbf{k}) = \frac{2}{e^{\beta\xi_{\mathbf{k}}} + 1} \tag{8.6}$$

The factor of two in f_0 is due to the spin degeneracy. The above equation defines the transport relaxation time $\tau_t(\mathbf{k})$. A more detailed derivation can be found in Ziman (1960). Here an equation for $\tau_t(\mathbf{k})$ is derived and solved. Then the distribution function is

$$f(\mathbf{k}) = f_0(\mathbf{k}) - e\tau_t(\mathbf{k})\mathbf{E} \cdot \nabla_k f(\mathbf{k}) \tag{8.7}$$

When the electric field is small, only a small amount of current flows. The system is only slightly out of equilibrium. The distribution function $f(\mathbf{k}) = f_0(\mathbf{k}) + f_1(\mathbf{k})$, where the change $f_1(\mathbf{k})$ is small. It is only necessary to retain terms of $O(E)$. Iterate the above equation, which

effectively replaces f by f_0 on the right-hand side. Then $f(\mathbf{k})$ is evaluated to $O(E)$. Furthermore, it is assumed that the system is isotropic and $f_0(\mathbf{k}) = f_0(k)$

$$f(\mathbf{k}) \approx f_0(k) - e\tau_t(\mathbf{k})\mathbf{E}\cdot\nabla_k f_0(k) \tag{8.8}$$

$$\approx f_0(k) - e\tau_t(\mathbf{k})\frac{\mathbf{E}\cdot\mathbf{k}}{m}\frac{df_0(k)}{d\varepsilon_k} \tag{8.9}$$

The electrical current density $\mathbf{J}$ is the product of the electron charge e, the electron's density n_0, and the average velocity $\langle\mathbf{v}\rangle$, which is obtained by averaging over the electron distribution:

$$\mathbf{J} = en_0\langle\mathbf{v}\rangle = e\int\frac{d^3k}{(2\pi)^3}f(\mathbf{k})\frac{\hbar\mathbf{k}}{m} \tag{8.10}$$

$$n_0 = \int\frac{d^3k}{(2\pi)^3}f(\mathbf{k}) \tag{8.11}$$

The distribution function $f(\mathbf{k})$ is normalized to give the electron density n_0. By using the result (8.9) for $f(\mathbf{k})$, the term f_0 gives an average $\langle\mathbf{v}\rangle$ of zero, since no current flows when there is no electric field; as many electrons are going one way as another. The current is proportional to the second term, and it is proportional to the electric field:

$$\mathbf{J} = e^2\int\frac{d^3k}{(2\pi)^3}\tau_t(k)\mathbf{v}_k(\mathbf{v}_k\cdot\mathbf{E})\left(-\frac{df_0(k)}{d\varepsilon_k}\right) \tag{8.12}$$

In a homogeneous, isotropic system, the current $\mathbf{J}$ flows in the direction of $\mathbf{E}$. The quantity $f_0(k)$ is independent of $\mathbf{k}$ direction. The only angular factors are $\mathbf{v}_k(\mathbf{v}_k\cdot\mathbf{E})$. The angular integrals will average this to $v_k^2\mathbf{E}/3$ in three dimensions. The conductivity σ is the ratio of $\mathbf{J}$ to $\mathbf{E}$:

$$\sigma = \frac{e^2}{3}\int\frac{d^3k}{(2\pi)^3}\tau_t(k)v_k^2\left(-\frac{df_0(k)}{d\varepsilon_k}\right) \tag{8.13}$$

It is a positive quantity since $df_0/d\varepsilon_k$ is always negative. Equation (8.13) is the basis of all the calculations. There remains the important task of deriving a formula for the relaxation time $\tau_t(k)$. It is *not* just the time between scattering events, which is derived from the imaginary part of the retarded self-energy. This distinction is important, since it makes life difficult. The relaxation time in the Boltzmann equation is a special quantity.

The impurities are assumed to be static, fixed, objects with a spherically symmetric potential. They have no internal excitations, so the electron scatters from them elastically. The impurity causes the electron in state $\mathbf{k}$ to scatter to $\mathbf{k}'$, which has the same energy, so that $|\mathbf{k}| = |\mathbf{k}'|$ and $\varepsilon_k = \varepsilon_{k'}$. The net rate of scattering out of the state $\mathbf{k}$ is the rate of going from $\mathbf{k}$ to $\mathbf{k}'$, which is proportional to $f(\mathbf{k})[1 - f(\mathbf{k}')]$ minus the rate from $\mathbf{k}'$ to $\mathbf{k}$, which is proportional to $f(\mathbf{k}')[1 - f(\mathbf{k})]$:

$$-\left(\frac{df}{dt}\right)_{\text{collisions}} = \frac{[f(\mathbf{k}) - f_0(\mathbf{k})]}{\tau_t(\mathbf{k})} = 2\pi n_i\int\frac{d^3k'}{(2\pi)^3}\delta(\varepsilon_k - \varepsilon_{k'})$$
$$\times\{|T_{\mathbf{k}\mathbf{k}'}|^2 f(\mathbf{k})[1 - f(\mathbf{k}')] - |T_{\mathbf{k}'\mathbf{k}}|^2 f(\mathbf{k}')[1 - f(\mathbf{k})]\} \tag{8.14}$$

Max - Planck - I
für Mikrostruktur
Bibliothek

where n_i is the concentration of impurities. The quantity $T_{\mathbf{kk}'}$, is the T-matrix element for scattering from $\mathbf{k}$ to $\mathbf{k}'$ which was defined in Sec. 4.1. It is symmetric in its indices $T_{\mathbf{kk}'} = T_{\mathbf{k}'\mathbf{k}}$, which simplifies the above equation

$$\frac{[f(\mathbf{k}) - f_0(\mathbf{k})]}{\tau_t(\mathbf{k})} = 2\pi n_i \int \frac{d^3k'}{(2\pi)^3} \delta(\varepsilon_k - \varepsilon_{k'}) |T_{\mathbf{kk}'}|^2 [f(\mathbf{k}) - f(\mathbf{k}')] \tag{8.15}$$

The integrand contains the factor $f(\mathbf{k}) - f(\mathbf{k}')$. The integral is evaluated by assuming the form in (8.9), which is written as

$$f(\mathbf{k}) = f_0(k) + \mathbf{k} \cdot \mathbf{E} C(k) \tag{8.16}$$

$$f(\mathbf{k}') = f_0(k) + \mathbf{k}' \cdot \mathbf{E} C(k) \tag{8.17}$$

$$f(\mathbf{k}) - f(\mathbf{k}') = (\mathbf{k} - \mathbf{k}') \cdot \mathbf{E} C(k) \tag{8.18}$$

Since $|\mathbf{k}| = |\mathbf{k}'|$ then and $f_0(k') = f_0(k)$. The quantities $f(\mathbf{k})$ and $f(\mathbf{k}')$ differ only in the angular part of the second term. The angular part is treated as follows: define a coordinate system in which the $\hat{z}$ direction is $\hat{k}$, so that

$$\hat{k} \cdot \hat{E} = \cos\theta \tag{8.19}$$

$$\hat{k} \cdot \hat{k}' = \cos\theta' \tag{8.20}$$

$$\hat{k}' \cdot \hat{E} = \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos\phi \tag{8.21}$$

where the law of cosines is used in the last identity. The difference of the two distribution functions is now

$$f(\mathbf{k}) - f(\mathbf{k}') = kEC(k)[\cos\theta(1 - \cos\theta') - \sin\theta \sin\theta' \cos\phi] \tag{8.22}$$

The last term on the right, which contains the factor $\cos\phi$, will vanish when doing the integral $\int d\phi$. There is no other ϕ dependence in the integrand of (8.15), and the average of $\cos\phi$ is zero. The remaining term may be written as

$$\int d\omega_{k'} [f(\mathbf{k}) - f(\mathbf{k}')] = kEC(k) \cos\theta \int d\Omega_{k'} (1 - \cos\theta') \tag{8.23}$$

$$= [f(\mathbf{k}) - f_0(k)] \int d\Omega_{k'} (1 - \cos\theta') \tag{8.24}$$

The term $f(\mathbf{k}) - f_0(k)$ is factored from both sides of (8.15), which leaves the definition for the reciprocal of the relaxation time:

$$\frac{1}{\tau_t(k)} = 2\pi n_i \int \frac{d^3k'}{(2\pi)^3} \delta(\varepsilon_k - \varepsilon_{k'}) |T_{\mathbf{kk}'}|^2 (1 - \cos\theta') \tag{8.25}$$

The important factor in the integrand is $(1 - \cos\theta') = 1 - \mathbf{k}\cdot\mathbf{k}'/k^2$. It makes the relaxation time in the Boltzmann equation different from the usual relaxation time $\tau(k)$, which is the time between scattering events. The latter quantity is simply

$$\frac{1}{\tau(k)} = 2\pi n_i \int \frac{d^3k'}{(2\pi)^3}\delta(\varepsilon_k - \varepsilon_{k'})|T_{\mathbf{kk}'}|^2 \tag{8.26}$$

$$= \frac{4\pi n_i}{mk}\sum_l (2l+1)\sin^2\delta_l(k) = v_k n_i \sigma(k) \tag{8.27}$$

$$\sigma(k) = \frac{4\pi}{k^2}\sum_l (2l+1)\sin^2\delta_l(k) \tag{8.28}$$

The usual inverse relaxation time is just the scattering cross section $\sigma(k)$ times the particle velocity v_k times the impurity density n_i. An equivalent result may be obtained for the relaxation time $\tau_t(k)$ of the Boltzmann equation. The T matrix is expanded as in (4.47) and (4.52) for the case where $|\mathbf{k}| = |\mathbf{k}'|$,

$$T_{\mathbf{kk}'} = -\frac{2\pi}{mk}\sum_l (2l+1)P_l(\cos\theta)e^{i\delta_l(k)}\sin(\delta_l(k)) \tag{8.29}$$

$$\frac{1}{\tau_t(k)} = \frac{4\pi n_i}{mk}\sum_{l=1} l\sin^2[\delta_l(k) - \delta_{l-1}(k)] \tag{8.30}$$

where θ is the angle between $\hat{k}$ and $\hat{k}'$. The angular integrals in (8.25) are straightforward but cumbersome, so only the result for τ_t is just listed. The formula for $1/\tau_t$ has a different combination of phase shifts than the formula for the cross section.

The factor $(1 - \cos\theta')$ weights the amount of scattering of the electron by the impurity. Small-angle scattering, where $\cos\theta' \approx 1$, is relatively unimportant in contributing to $1/\tau_t$. These events do little to impede the flow of electrons and so contribute little to the resistivity. The factor $(1 - \cos\theta')$ obviously favors large-angle scattering events, which are more important for the electrical resistivity. The relaxation time in the transport equation is not identical to the average scattering rate because there is an additional factor to weight the amount of scattering.

An example of evaluating the conductivity in (8.13) is given for a free-electron metal. At low temperature the distribution function (8.6) becomes a delta function in energy which sets $k = k_F$

$$\left(-\frac{df_0}{d\xi_k}\right) = 2\delta(\xi_k) \tag{8.31}$$

The angular integrals have already been done, so that

$$\sigma = \frac{2e^2}{3}\frac{4\pi}{(2\pi)^3}v_F^2 m k_F \tau_t(k_F) = \frac{e^2 n_0 \tau_t(k_F)}{m} \tag{8.32}$$

where $n_0 = k_F^3/3\pi^2$. The relevant relaxation time is for electrons at the Fermi surface. Yet the conductivity is proportional to the density n_0 of all conduction electrons and not just to those at the Fermi surface. This surprising result is quite reasonable once the physics is understood. When the electric field is first imposed, the equation $\mathbf{k} = -e\mathbf{E}$ shows that all electrons in the Fermi sea start accelerating equally. The Fermi sea is translationally shifted in wave vector space. The scattering tends to relax the Fermi distribution back to its undisturbed configuration. As shown in Fig. 8.1, electrons in the leading edge of the displaced Fermi distribution

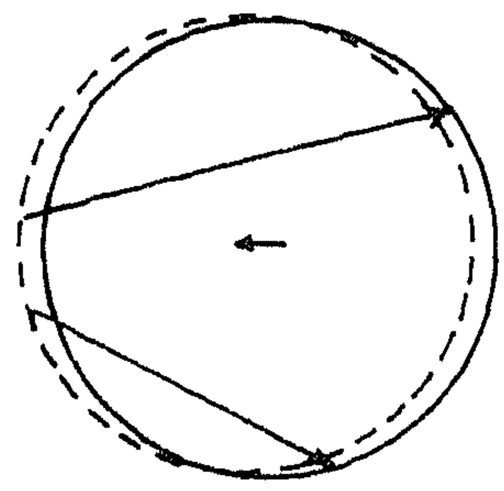


FIGURE 8.1 The circle represents the Fermi sea, which begins to move in response to an applied electric field. Steady state is maintained by relaxation back to other points on the Fermi surface.

are scattered back to the rear regions. Only those electrons at the Fermi surface can scatter. The electrons well below the Fermi surface cannot elastically scatter, since all states with the same energy are already occupied with other electrons. Above the Fermi surface there are no thermally excited electrons. Only electrons at the Fermi surface are available to elastically scatter to other points on the Fermi surface. The lifetime only involves $\tau_t(k_F)$.

The conductivity is relatively insensitive to temperature as long as the density of states of the metal is a smooth function of energy near the Fermi surface. The resistivity $\rho = 1/\sigma$ is the inverse of the conductivity, and is proportional to the concentration of impurities. This proportionality is experimentally verified.

Since impurity scattering is elastic, it does not change the energy of the electron. As the current of electrons moves through the solid, each electron gains energy from the electric field. How does the electron lose this energy, if it only scatters elastically? The next few paragraphs will answer this question.

Let $\nu = \cos\theta$, where θ is the angle between $\mathbf{k}$ and $\mathbf{E}$. In order to keep the discussion simple, it is assumed that the distribution $f(\mathbf{k})$ is isotropic in the absence of a field. When the field is present, the distribution function can be expanded in a Legendre series in ν. The first few terms are

$$f(\mathbf{k}) = f_s + \nu f_p(k) + P_2(\nu) f_d(k) + \cdots \tag{8.33}$$

where $f_s(k)$ is the isotropic part of the distribution while $f_p(k)$ is the $l = 1$ part of the distribution. Note that $f_s(k)$ is not the equilibrium part of the distribution, which is $f_0(k)$. The electrical current is determined by the distribution f_p.

The conduction process can be viewed as having the following steps:

1. For $t < 0$, $\mathbf{E} = 0$ and the initial distribution is f_0.
2. At $t = 0$ the field $\mathbf{E}$ is switched on. It accelerates the particles and creates the distribution f_p.
3. The elastic scattering takes the particles from the p distribution f_p to the s distribution f_s. This step has a time constant τ_t.
4. The isotropic distribution f_s relaxes back to the equilibrium distribution f_0. This step has a different time constant.

The energy relaxation occurs in the last step where f_s is brought to equilibrium. The electrons can lose energy to their heat bath, usually phonons. This process has a very different time constant than τ_t and is usually much slower. The relaxation time τ_t from elastic scattering determines the rate at which particles scatter out of the p distribution f_p into other distributions such as f_s and f_d. The current is determined by τ_t since it gives the steady state amplitude of f_p.

The energy relaxation occurs elsewhere in the chain of events. The rate for energy relaxation τ_e is calculated below.

8.1.2. Kubo Formula: Approximate Solution

The electrical conductivity can be calculated from the Kubo formula by using the technique described in (3.390)–(3.393). The correlation function is evaluated for nonzero temperatures and frequencies:

$$\pi(i\omega_n) = -\frac{1}{3\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \mathbf{j}(\tau)\cdot\mathbf{j}(0)\rangle \tag{8.34}$$

The retarded function $\pi_{\text{ret}}(\omega)$ is obtained by letting $i\omega_n \to \omega + i\delta$, and the dc conductivity is given by the limit of $\omega \to 0$:

$$\sigma = -\lim_{\omega\to 0}\left\{\frac{\text{Im}[\pi_{\text{ret}}(\omega)]}{\omega}\right\} \tag{8.35}$$

This correlation function for σ is evaluated for the same model system described in Sec. 8.1.1. There is a free-particle system with a dilute concentration of simple scattering centers. Equation (8.34) is evaluated for the following Hamiltonian and current operator:

$$H = \sum_{\mathbf{p}\sigma}\xi_p C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} + V \tag{8.36}$$

$$V = \frac{1}{\nu}\sum_{\mathbf{q}} V(q)\rho_i(\mathbf{q})\rho(\mathbf{q}) \tag{8.37}$$

$$\rho_i(\mathbf{q}) = \sum_j e^{i\mathbf{q}\cdot\mathbf{R}_j} \tag{8.38}$$

$$\rho(\mathbf{q}) = \sum_{\mathbf{p}\sigma} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} \tag{8.39}$$

$$\mathbf{j} = \frac{e}{m}\sum_{\mathbf{p}\sigma}\mathbf{p}C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} \tag{8.40}$$

The impurities are at positions $\mathbf{R}_j$, and an average will be taken over the possible distributions of impurity positions. This averaging technique was described earlier in Sec. 4.1.4.

The theoretical calculation is divided into two parts. The first part, in this section, is simply to reproduce the Boltzmann result which was derived above. How is a conductivity derived which is proportional to the relaxation time τ_t? The derivation entails a summation over a set of vertex diagrams. The treatment is kept as introductory as possible, since it is one of the first summations over vertex diagrams. A formally exact solution to the correlation function due to Langer (1960) is presented in Sec. 8.1.3.

The logical way to evaluate the correlation function (8.34) is as a power series in the concentration of impurities. Averages over the impurity density operators in (4.91) are expressed as a power series in the number of impurities N_i:

$$\begin{aligned} f_n(\mathbf{q}_1,\ldots,\mathbf{q}_n) &= \langle\rho_i(\mathbf{q}_1)\rho_i(\mathbf{q}_2)\cdots\rho_i(\mathbf{q}_n)\rangle \\ &= N_i\delta_{\Sigma\mathbf{q}_j} + N_i^2\delta_{\Sigma\mathbf{q}_i}\delta_{\Sigma_j\mathbf{q}_j} + \cdots \end{aligned} \tag{8.41}$$

At first sight it appears possible to evaluate (8.34) by just expanding the S matrix and collecting all terms proportional to N_i, then N_i^2, etc. A simple expansion in powers of N_i does

not work, as is apparent from the desired answer. It has $\sigma \propto \tau_t \propto 1/n_i$. The first term has the conductivity inversely proportional to n_i, which is impossible to obtain as a simple series in N_i except by summing a set of diagrams. The expected answer is only going to be obtained by summing a series of diagrams.

The correlation function (8.34) is written in the interaction representation as

$$\pi(i\omega_n) = -\frac{1}{3\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau S(\beta)e^{\tau H_0}\mathbf{j}e^{-\tau H_0}\cdot\mathbf{j}\rangle \tag{8.42}$$

The logical way to evaluate this expression is to expand the S matrix and consider each term. The first term has $S = 1$, and this correlation function is called π_0:

$$\begin{aligned}\pi_0(i\omega_n) &= -\frac{e^2}{3m^2\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\sum_{\mathbf{p}\mathbf{p}'\sigma\sigma'}\mathbf{p}\cdot\mathbf{p}'\langle T_\tau C^\dagger_{\mathbf{p}\sigma}(\tau)C_{\mathbf{p}\sigma}(\tau)C^\dagger_{\mathbf{p}'\sigma'}(0)C_{\mathbf{p}'\sigma'}(0)\rangle \\ &= \frac{2e^2}{3m^2\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\sum_{\mathbf{p}}p^2\int_0^\beta d\tau e^{i\omega_n\tau}\mathscr{G}^{(0)}(\mathbf{p},\tau)\mathscr{G}^{(0)}(\mathbf{p},-\tau)\end{aligned} \tag{8.43}$$

This expression is zero unless $i\omega_n = 0$, since the number operators $C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}$ are τ-independent. The τ-integral gives zero unless $i\omega_n = 0$ and then it gives β. The term π_0 gives a conductivity of zero. The result is not surprising, since it is the correlation function of noninteracting particles, and they have zero resistivity. Perhaps a better answer is the correlation function for the conductivity is infinity. The zero is sufficient to alert us that a nonsensical question was asked, and a nonsensical answer was obtained. A resistive system requires putting damping into the particle motion.

The next logical step is to replace all $\mathscr{G}^{(0)}$ by $\mathscr{G}$. The self-energy of the particles, from impurity scattering, is included in all the particle Green's functions. Of course, $\mathscr{G}$ is obtained by summing a series of diagrams, which is Dyson's equation. The self-energy $\Sigma(\mathbf{p}, ip_n)$ from impurity scattering was evaluated in Sec. 4.1 in the limit of low n_i. It is a retarded function with real and imaginary parts, where the imaginary parts are due to the damping of the particle motion. The step of replacing $\mathscr{G}^{(0)}$ by $\mathscr{G}$ does put in damping of the particle motion. The first correlation function which will be evaluated is shown in Fig. 8.2. It is a simple bubble diagram, with the smooth lines denoting $\mathscr{G}$ and the two vertices having the vector vertex $\mathbf{p}$. This correlation function is called $\pi^{(0)}(i\omega)$:

$$\pi^{(0)}(i\omega_n) = \frac{2e^2}{3m^2\nu}\sum_{\mathbf{p}}p^2\frac{1}{\beta}\sum_{ip}\mathscr{G}(\mathbf{p}, ip + i\omega)\mathscr{G}(\mathbf{p}, ip) \tag{8.44}$$

The wiggly lines at the two ends of the bubble, which are connected to the vertices, represent the incoming frequency $i\omega_n$. The first step is to evaluate the summation over Matsubara frequencies.

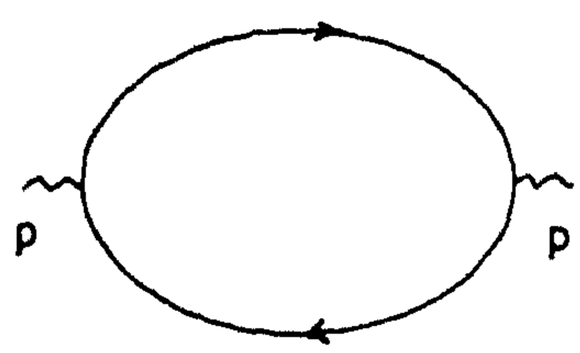


FIGURE 8.2

The procedure for doing this summation was described in Sec. 3.5. The easiest way is to represent the interacting Green's functions by the Lehmann representation

$$S(i\omega_n) = \frac{1}{\beta}\sum_{ip} \mathscr{G}(\mathbf{p}, ip + i\omega)\mathscr{G}(\mathbf{p}, ip)$$

$$= \int \frac{d\varepsilon_1}{2\pi} A(\mathbf{p}, \varepsilon_1) \int \frac{d\varepsilon_2}{2\pi} A(\mathbf{p}, \varepsilon_2) \frac{1}{\beta}\sum_{ip_m} \frac{1}{ip_m + i\omega_n - \varepsilon_1} \frac{1}{ip_m - \varepsilon_2} \tag{8.45}$$

$$= \int \frac{d\varepsilon_1}{2\pi} A(\mathbf{p}, \varepsilon_1) \int \frac{d\varepsilon_2}{2\pi} A(\mathbf{p}, \varepsilon_2) \frac{n_F(\varepsilon_2) - n_F(\varepsilon_1)}{i\omega_n + \varepsilon_2 - \varepsilon_1} \tag{8.46}$$

The next step is to convert to a retarded function ($i\omega_n \to \omega + i\delta$) and then to take the imaginary part:

$$\mathrm{Im}[S(\omega)] = -\pi \int \frac{d\varepsilon_1}{2\pi} A(\mathbf{p}, \varepsilon_1) \int \frac{d\varepsilon_2}{2\pi} A(\mathbf{p}, \varepsilon_2)\{n_F(\varepsilon_2) - n_F(\varepsilon_1)\}\delta(\omega + \varepsilon_2 - \varepsilon_1)$$

$$= -\int \frac{d\varepsilon_2}{4\pi} A(\mathbf{p}, \varepsilon_2) A(\mathbf{p}, \varepsilon_2 + \omega)\{n_F(\varepsilon_2) - n_F(\varepsilon_2 + \omega)\} \tag{8.47}$$

$$\mathrm{Im}[\pi^{(0)}_{\mathrm{ret}}(\omega)] = -\frac{e^2}{3m^2 \mathrm{v}} \sum_{\mathbf{p}} p^2 \int \frac{d\varepsilon_2}{2\pi} A(\mathbf{p}, \varepsilon_2) A(\mathbf{p}, \varepsilon_2 + \omega)\{n_F(\varepsilon_2) - n_F(\varepsilon_2 + \omega)\}$$

The next step in the derivation (8.35) is to divide by ω and then to take the limit $\omega \to 0$. The important frequency dependence is in the last factor:

$$\lim_{\omega \to 0} \frac{1}{\omega}\{n_F(\varepsilon_2) - n_F(\varepsilon_2 + \omega)\} = -\frac{dn_F(\varepsilon_2)}{d\varepsilon_2} \tag{8.48}$$

and the conductivity from this contribution is called $\sigma^{(0)}$

$$\sigma^{(0)} = \frac{e^2}{3m^2} \int \frac{d^3p}{(2\pi)^3} p^2 \int \frac{d\varepsilon}{2\pi} A^2(\mathbf{p}, \varepsilon)\left(-\frac{dn_F(\varepsilon)}{d\varepsilon}\right) \tag{8.49}$$

In the last step, the limit $\mathrm{v} \to \infty$ changed the summation over $\mathbf{p}$ into a continuous integral. The right-hand side is positive since $dn_F/d\varepsilon$ is negative. Before discussing this result, it is useful to review the order of the steps in the derivation. They will be used in all evaluations of the Kubo formula:

1. Do all summations over Matsubara frequencies ip_n.
2. Analytically continue $i\omega \to \omega + i\delta$ to get the retarded function $\pi_{\mathrm{ret}}(\omega)$.
3. Take the imaginary part $\mathrm{Im}[\pi_{\mathrm{ret}}(\omega)]$.
4. Divide by ω, and then take the limit $\omega \to 0$.

These steps cannot be taken out of order. Equation (8.49) for $\sigma^{(0)}$ has several interesting features. The factor $dn_F/d\varepsilon = -\delta(\varepsilon)$ at zero temperature, which is rather convenient, since it serves to eliminate the integral over $d\varepsilon$. The sharp step in $n_F(\varepsilon)$ is in contrast to the momentum distribution (3.135),

$$n_{\mathbf{p}} = \int \frac{d\varepsilon}{2\pi} n_F(\varepsilon) A(\mathbf{p}, \varepsilon) \tag{8.50}$$

which no longer has a discontinuous step at $p = k_F$ since the impurity scattering causes a general smearing of this distribution. However, the energy distribution $n_F(\varepsilon)$ is always a sharp function of ε at zero temperature regardless of the interactions.

Another crucial feature of (8.49) is that the spectral function is squared, $A(\mathbf{p}, \varepsilon)^2$. That this is important may be shown by examining the limit as the impurity concentration n_i vanishes. Since the self-energy is proportional to n_i, it will vanish in this limit. Define $\Delta = -\mathrm{Im}(\Sigma)$, and then

$$\lim_{n_i \to 0} A(\mathbf{p}, \varepsilon) = \lim_{\Delta \to 0} \frac{2\Delta}{[\varepsilon - \xi_{\mathbf{p}} - \mathrm{Re}(\Sigma)]^2 + \Delta^2} = 2\pi\delta(\varepsilon - \xi_{\mathbf{p}}) \tag{8.51}$$

The spectral function becomes a delta function when $n_i \to 0$. This limiting behavior is reasonable, since in the absence of self-energy effects, the particles are all free, and the spectral function is indeed a delta function. The question at hand is what happens to A^2 as $n_i \to 0$, since it appears to go as the square of a delta function. In fact, A^2 does diverge as $n_i \to 0$, which makes the conductivity diverge to infinity when $n_i \to 0$. A method is needed for handling this divergence. The answer is provided by considering the integrals

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{2\Delta}{\omega^2 + \Delta^2} = 1 \tag{8.52}$$

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left(\frac{2\Delta}{\omega^2 + \Delta^2} \right)^2 = \frac{1}{\Delta} \tag{8.53}$$

where $\omega = \varepsilon - \xi_{\mathbf{p}} - \mathrm{Re}(\Sigma)$. The first integral has the right behavior as $\Delta \to 0$, since it gives the same result as does $A = 2\Delta/(\omega^2 + \Delta^2) = 2\pi\delta(\omega)$. The second integral suggests the replacement

$$\lim_{\Delta \to 0} A^2 = \lim_{\Delta \to 0} \left(\frac{2\Delta}{\omega^2 + \Delta^2} \right)^2 = \lim_{\Delta \to 0} \frac{2\pi\delta(\omega)}{\Delta} \tag{8.54}$$

which will give the right behavior as $\Delta \to 0$. The replacement $A^2 \to 2\pi\delta(\omega)/\Delta$ will be made in the limit as $n_i \to 0$. Furthermore, the quantity 2Δ is recognized as the inverse scattering time of electrons on the Fermi surface,

$$2\Delta(k_F, \varepsilon = 0) = \frac{1}{\tau(k_F)} = -2\ \mathrm{Im}(\Sigma) \tag{8.55}$$

The conductivity formula may now be written as

$$\sigma^{(0)} = \frac{2e^2}{3m^2} \int \frac{d^3p}{(2\pi)^3} \delta(\xi_p) p^2 \tau(p) \tag{8.56}$$

This equation looks like the right answer for σ since it seems to have exactly the same combination of factors as (8.13). But there is a very important difference between (8.56) and the Boltzmann result (8.13)—in the relaxation time. The formula (8.56) has a relaxation time without the $(1 - \cos\theta')$ factor, since the relaxation time in (8.56) is from $\mathrm{Im}(\Sigma)$, which is just the average time between scattering events. The $(1 - \cos\theta')$ factor was important for weighting the large-angle scattering processes, which were important for the resistivity. The preliminary answer (8.56) is not the Boltzmann result and has serious deficiencies.

The above derivation has one achievement. It succeeds in deriving a term in σ which is inversely proportional to n_i. The relaxation time $\tau(p)$, although the wrong one for the

conductivity, at least has the virtue that it is inversely proportional to n_i, which makes $\sigma^{(0)}$ also inversely proportional to n_i.

The calculation has not yet yielded the right result. More diagrams need to be evaluated. Among those remaining, there must be a subset which, when evaluated, will give the right answer. The S matrix is expanded for impurity scattering, and the remaining terms contain higher powers in the impurity interaction V. At first it appears that higher powers in the impurity interaction must imply that the additional terms are higher powers in the impurity concentration n_i. This conclusion is false, as is obvious from the answer. The final σ is proportional to $1/n_i$, while our preliminary term $\sigma^{(0)}$ is also proportional to $1/n_i$. The important correction terms in the S-matrix expansion must also yield terms in the conductivity which are proportional to $1/n_i$. The S-matrix expansion is examined to find the terms which cause σ to diverge as $1/n_i$ when $n_i \to 0$, although these terms must come from higher terms in the S-matrix expansion. Higher-order terms in S can be proportional to $1/n_i$ if they also contain higher powers of the spectral function A^n.

The correlation function $\sigma^{(0)}$, contained Green's functions $\mathscr{G}$ which include all self-energy effects. The remaining diagrams are called *vertex corrections*. They are defined as diagrams in which the impurity scattering links the Green's functions on both sides of the bubble. Some vertex diagrams are shown in Fig. 8.3(a). There is a single impurity with a varying number of scattering events from the electron line on either side of the bubble. If there were no scattering line connecting one side of the bubble, the diagram would be a self-energy term on the other side. A diagram in which the two electron lines, on both sides of the bubble, scatter from the same impurity cannot be a self-energy diagram of either one and so is called a vertex diagram. Figure 8.3(a) only shows vertex diagrams with a single impurity participating in the scattering. Vertex diagrams can occur with scattering from several impurities. These are equally important and are considered later.

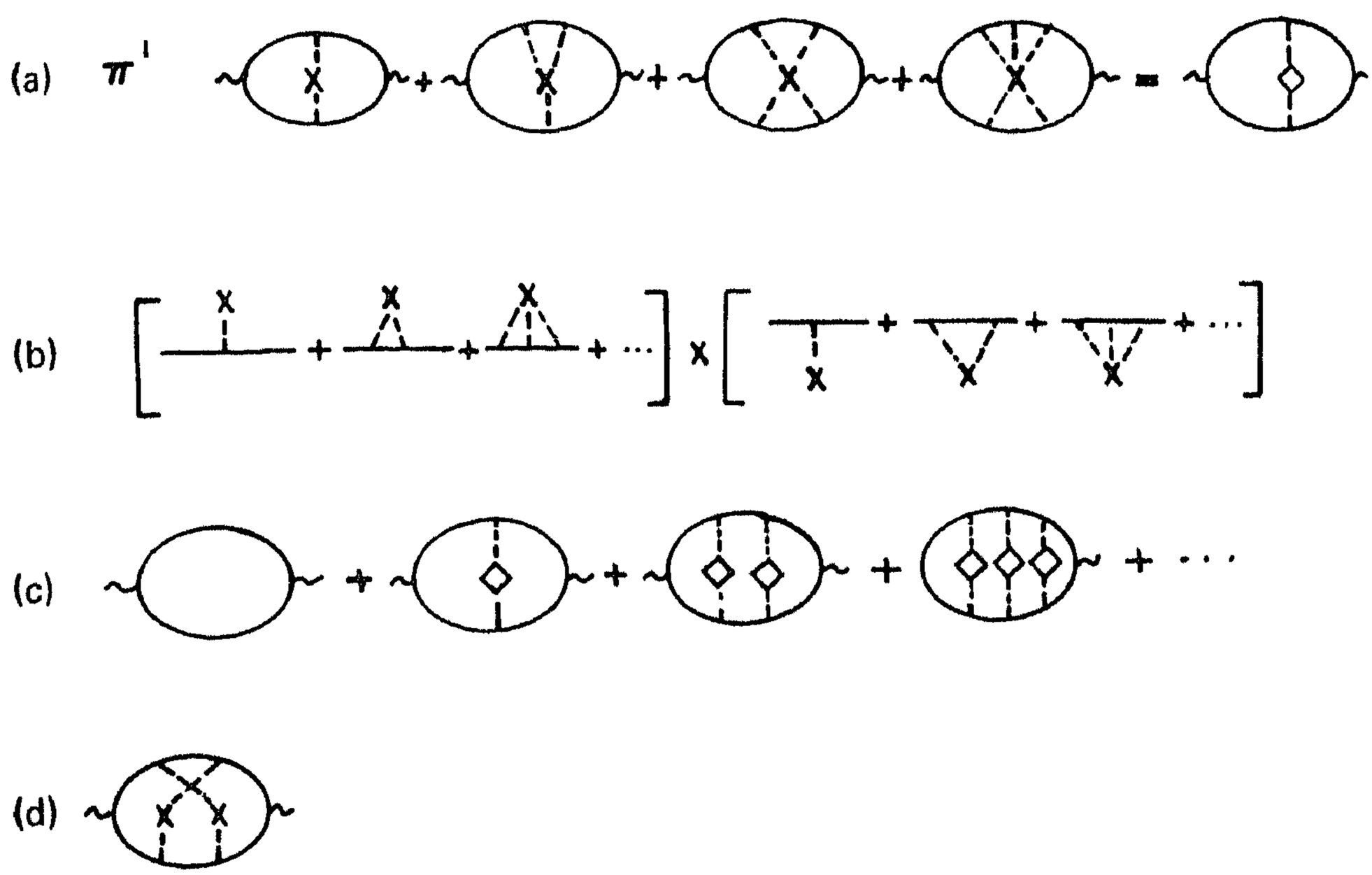


FIGURE 8.3

The sum of all the diagrams in Fig. 8.3(a) can be evaluated in a simple way and gives the correlation function:

$$W^{(1)}_{\mathbf{p}\mathbf{p}'}(ip, ip + i\omega) = n_i T_{\mathbf{p}\mathbf{p}'}(ip) T_{\mathbf{p}'\mathbf{p}}(ip + i\omega) \tag{8.57}$$

$$\pi^{(1)}(i\omega) = \frac{2e^2}{3m^2\nu^2} \sum_{\mathbf{p}\mathbf{p}'} \frac{1}{\beta} \sum_{ip} \mathbf{p} \cdot \mathbf{p}' \mathscr{G}(\mathbf{p}, ip) \mathscr{G}(\mathbf{p}, ip + i\omega) \mathscr{G}(\mathbf{p}', ip)$$
$$\times \mathscr{G}(\mathbf{p}', ip + i\omega) W^{(1)}_{\mathbf{p}\mathbf{p}'}(ip, ip + i\omega) \tag{8.58}$$

Equation (8.58) is easy to prove once the rules are recalled for scattering from a single impurity: (1) total momentum is conserved, and (2) energy is not changed from ip or $ip + i\omega$. The second rule says that all electron lines in one string, say the top, are at the same energy ip, while the ones on the bottom are at $ip + i\omega$. The momentum conservation requires that the momentum transfer on the top electron line be $\mathbf{p} - \mathbf{p}'$, which is exactly the opposite of the other: $\mathbf{p}' - \mathbf{p}$. The evaluation is easy, since the two sides decouple, as is illustrated schematically in Fig. 8.3(b). The effective interaction $W^{(1)}$ is represented by a diamond in the figures, which is illustrated to the right of Fig. 8.3(a). The diamond is the total vertex scattering from a single impurity.

Figure 8.3(c) shows the sum of correlation functions which have increasing numbers of diamonds in them. These contributions are called *ladder diagrams*. Summing the ladder diagrams achieves the objective of a relaxation time with the factor $(1 - \cos\theta')$. It is important to realize that the sum of diagrams in Fig. 8.3(c) is not the only contribution with scattering from several impurity sites. An example of a nonladder diagram is shown in Fig. 8.3(d). This contribution is not included in the series shown in Fig. 8.3(c). It is of order $O(n_i^2)$ and is neglected. The sum of ladder diagrams omits many terms. However, the omitted terms are not as important in the limit where $n_i \to 0$.

The first two terms in the sum of ladder diagrams have already been derived; they are $\pi^{(0)}$, and $\pi^{(1)}$, in (8.44) and (8.58). The superscript denotes the number of ladders. The superscript (L) denotes sum of ladder diagrams. The series of terms in the ladder sum can be generated by representing them as a vector vertex function $\mathbf{\Gamma}^{(L)}$:

$$\pi^{(L)}(i\omega) = \frac{2e^2}{3m^2\nu} \sum_{\mathbf{p}} \frac{1}{\beta} \sum_{ip} \mathbf{p} \cdot \mathbf{\Gamma}^{(L)}(\mathbf{p}, ip, ip + i\omega) \mathscr{G}(\mathbf{p}, ip) \mathscr{G}(\mathbf{p}, ip + i\omega) \tag{8.59}$$

$$\mathbf{\Gamma}^{(L)}(\mathbf{p}; ip, ip + i\omega) = \mathbf{p} + \frac{1}{\nu} \sum_{\mathbf{p}'} \mathbf{\Gamma}^{(L)}(\mathbf{p}'; ip, ip + i\omega) W^{(1)}_{\mathbf{p}\mathbf{p}'}(ip, ip + i\omega)$$
$$\times \mathscr{G}(\mathbf{p}', ip) \mathscr{G}(\mathbf{p}', ip + i\omega) \tag{8.60}$$

Repeated iteration of (8.60) will generate the series of terms in the ladder summation shown in Fig. 8.3(c). The ladder summation will be evaluated, approximately, in order to obtain the factor of $(1 - \cos\theta')$ in the relaxation time.

The correlation function is a function of ip and $i\omega$ in the combination ip and $ip + i\omega$. Define the quantity

$$\begin{aligned} P(ip, ip + i\omega) &= \frac{2e^2}{3m^2\nu} \sum_{\mathbf{p}} \mathbf{p} \cdot \mathbf{\Gamma}^{(L)}(\mathbf{p}; ip, ip + i\omega) \mathscr{G}(\mathbf{p}, ip) \mathscr{G}(\mathbf{p}, ip + i\omega) \\ \pi^{(L)}(i\omega) &= \frac{1}{\beta} \sum_{ip} P(ip, ip + i\omega) \end{aligned} \tag{8.61}$$

The summation over Matsubara frequencies ip is evaluated, as usual, by examining the contour integral $\int dz n_F(z) P(z, z + i\omega)$. The integrand has the poles of $n_F(z)$, which give the summation over ip, and also branch cuts along the two axes $z = \varepsilon$ and $z = \varepsilon - i\omega$, where ε is real. These cuts are shown in Fig. 8.4. The integral equals the contribution from the two branch cuts, where one has to subtract the parts above and below each cut:

$$\pi^{(L)}(i\omega_n) = -\int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi i} n_F(\varepsilon)[P(\varepsilon + i\delta, \varepsilon + i\omega_n) - P(\varepsilon - i\delta, \varepsilon + i\omega_n) + P(\varepsilon - i\omega_n, \varepsilon + i\delta) - P(\varepsilon - i\omega_n, \varepsilon - i\delta)] \tag{8.62}$$

The infinitesimal part $\pm i\delta$ is unnecessary when there is a term with $\pm i\omega_n$. Next find the retarded function from the analytical continuation $i\omega_n \to \omega + i\delta$. A variable change $\varepsilon \to \varepsilon + \omega$ in the last two terms brings us to the point

$$\pi_{\text{ret}}^{(L)}(\omega) = \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi i} \{[n_F(\varepsilon) - n_F(\varepsilon + \omega)]P(\varepsilon - i\delta, \varepsilon + \omega + i\delta) - n_F(\varepsilon)P(\varepsilon + i\delta, \varepsilon + \omega + i\delta) + n_F(\varepsilon + \omega)P(\varepsilon - i\delta, \varepsilon + \omega - i\delta)\} \tag{8.63}$$

The next step is to take the imaginary part of this expression. Because of the factor of i in the denominator $2\pi i$, this step gives the real part of the integrand. The subsequent step is to take the limit that $\omega \to 0$. The function $P(\varepsilon - i\delta, \varepsilon + \omega + i\delta)$ becomes $P(\varepsilon - i\delta, \varepsilon + i\delta)$ at zero frequency, which is real. It is real because the function is symmetric in its arguments, $P(z_1, z_2) = P(z_2, z_1)$, so that the complex conjugate of $P(\varepsilon - i\delta, \varepsilon + i\delta)$ equals itself. The last two factors of P in the above equation are complex conjugates of each other. Taking the real part of the integrand removes their imaginary parts, and the real parts are equal.

The limit $\omega \to 0$ again only involves $[n_F(\varepsilon) - n_F(\varepsilon + \omega)]/\omega$. These steps brings us to a formal expression for the dc conductivity:

$$\sigma = -\lim_{\omega \to 0}\left(\frac{\text{Im}[\pi_{\text{ret}}^{(L)}]}{\omega}\right) = \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi i}\left(-\frac{dn_F(\varepsilon)}{d\varepsilon}\right)\{P(\varepsilon - i\delta, \varepsilon + i\delta) - \text{Re}[P(\varepsilon + i\delta, \varepsilon + i\delta)]\}$$

There are only two functions which need to be found, $P(\varepsilon - i\delta, \varepsilon + i\delta)$ and $P(\varepsilon + i\delta, \varepsilon + i\delta)$. At zero temperature, where $-dn_F/d\varepsilon = \delta(\varepsilon)$, they need to be found only at $\varepsilon = 0$. These two functions have quite different behavior and are obtained by different methods. Both are usually important, but the most singular is $P(\varepsilon - i\delta, \varepsilon + i\delta)$, and this term leads to the factor of $(1 - \cos\theta')$ in the lifetime.

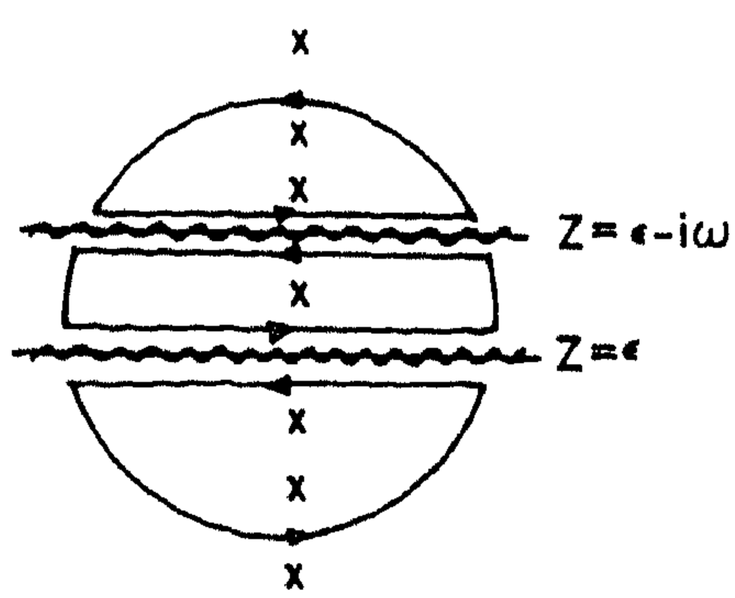


FIGURE 8.4 Branch cuts in the contour integral.

Consider the definition of P in (8.61):

$$P(\varepsilon - i\delta, \varepsilon + i\delta) = \frac{2e^2}{3m^2 \mathrm{v}} \sum_{\mathbf{p}} \mathbf{p} \cdot \mathbf{\Gamma}^{(L)}(\mathbf{p}, \varepsilon - i\delta, \varepsilon + i\delta) \mathscr{G}(\mathbf{p}, \varepsilon - i\delta) \mathscr{G}(\mathbf{p}, \varepsilon + i\delta)$$

The product of Green's functions is entirely real:

$$\begin{aligned} \mathscr{G}(\mathbf{p}, \varepsilon - i\delta)\mathscr{G}(\mathbf{p}, \varepsilon + i\delta) &= G_{\mathrm{adv}}(\mathbf{p}\varepsilon) G_{\mathrm{ret}}(\mathbf{p}, \varepsilon) = \frac{1}{\omega^2 + \Delta^2} \\ &= \frac{A(\mathbf{p}, \varepsilon)}{2\Delta(\mathbf{p}, \varepsilon)} \end{aligned} \tag{8.65}$$

$$\begin{aligned} \omega &= \varepsilon - \xi_{\mathbf{p}} - \mathrm{Re}[\Sigma(\mathbf{p}, \varepsilon)], \qquad \Delta = -\mathrm{Im}[\Sigma] \\ P(\varepsilon - i\delta, \varepsilon + i\delta) &= \frac{e^2}{3m^2} \int \frac{d^3p}{(2\pi)^3} \frac{A(\mathbf{p}, \varepsilon)}{\Delta(\mathbf{p}, \varepsilon)} \mathbf{p} \cdot \mathbf{\Gamma}^{(L)}(\mathbf{p}, \varepsilon - i\delta, \varepsilon + i\delta) \end{aligned} \tag{8.66}$$

The combination $G_{\mathrm{adv}} G_{\mathrm{ret}}$ is rigorously defined as the spectral function $A(\mathbf{p}, \varepsilon)$ divided by $2\Delta(\mathbf{p}, \varepsilon)$. There is no assumption that $A^2 = A/2\Delta$. The same combination is found in the equation for the vertex function:

$$\begin{aligned} \mathbf{\Gamma}^{(L)}(\mathbf{p}, \varepsilon - i\delta, \varepsilon + i\delta) = \mathbf{p} + \int \frac{d^3p'}{(2\pi)^3} \mathbf{\Gamma}^{(L)}(\mathbf{p}', \varepsilon - i\delta, \varepsilon + i\delta) \\ \times W^{(1)}_{\mathbf{pp}'}(\varepsilon - i\delta, \varepsilon + i\delta) G_{\mathrm{adv}}(\mathbf{p}', \varepsilon) G_{\mathrm{ret}}(\mathbf{p}', \varepsilon) \end{aligned} \tag{8.67}$$

The vector function $\mathbf{\Gamma}$ must point in the direction of $\mathbf{p}$ since that is the only vector in its function arguments. It is convenient to define an integral equation for the scalar function:

$$\mathbf{p}\gamma(\mathbf{p}, \varepsilon) = \mathbf{\Gamma}^{(L)}(\mathbf{p}, \varepsilon - i\delta, \varepsilon + i\delta) \tag{8.68}$$

$$\gamma(\mathbf{p}, \varepsilon) = 1 + \int \frac{d^3p'}{(2\pi)^3} \frac{A(\mathbf{p}', \varepsilon)}{2\Delta(\mathbf{p}', \varepsilon)} \frac{\mathbf{p} \cdot \mathbf{p}'}{p^2} W^{(1)}_{\mathbf{pp}'}(\varepsilon - i\delta, \varepsilon + i\delta) \gamma(\mathbf{p}', \varepsilon) \tag{8.69}$$

$$P(\varepsilon - i\delta, \varepsilon + i\delta) = \frac{e^2}{3m^2} \int \frac{d^3p}{(2\pi)^3} \frac{A(\mathbf{p}, \varepsilon)}{\Delta(\mathbf{p}, \varepsilon)} p^2 \gamma(\mathbf{p}, \varepsilon) \tag{8.70}$$

Equation (8.69) is a one-dimensional integral equation for the scalar function $\gamma(\mathbf{p}, \varepsilon)$, where p' is the integration variable. The angular integrals $d\Omega_{p'}$ just average the quantity $\mathbf{p} \cdot \mathbf{p}' W^{(L)}_{\mathbf{pp}'}$ over angles to provide the kernel for the integral equation. The integral equation should not be difficult to solve with modern computers for realistic self-energies and T-matrices.

The Boltzmann result is obtained in the twin limits $T \to 0$ ($\varepsilon \to 0$) and $n_i \to 0$, where $A(\mathbf{p}', 0) \to 2\pi\delta(\xi_{\mathbf{p}'})$. Equation (8.69) then reduces to the integral equation

$$\gamma(\mathbf{p}, \varepsilon) = 1 + \int \frac{d^3p'}{(2\pi)^3} \frac{2\pi\delta(\xi_{\mathbf{p}'})}{2\Delta(\mathbf{p}', \varepsilon)} \frac{\mathbf{p} \cdot \mathbf{p}'}{p^2} n_i |T_{\mathbf{pp}'}|^2 \gamma(\mathbf{p}', \varepsilon) \tag{8.71}$$

$$\gamma(k_F) = 1 + \gamma(k_F) \frac{\Delta - \Delta_1}{\Delta} \tag{8.72}$$

$$2\Delta_1 = 2\pi n_i \int \frac{d^3p'}{(2\pi)^3} \delta(\xi_{\mathbf{p}'}) \left[1 - \frac{\mathbf{p} \cdot \mathbf{p}'}{p^2} \right] |T_{\mathbf{pp}'}|^2 = \frac{1}{\tau_t} \tag{8.73}$$

which is easily solved to give

$$\gamma(k_F) = \frac{\Delta}{\Delta_1} \tag{8.74}$$

The factor $(1 - \mathbf{p}\cdot\mathbf{p}'/p^2) = (1 - \cos\theta')$ since $|\mathbf{p}'| = |\mathbf{p}|$. The solution $\gamma = \Delta/\Delta_1$ is put into (8.65) to give

$$P(\varepsilon - i\delta, \varepsilon + i\delta) = \frac{e^2}{3m^2}\int \frac{d^3p}{(2\pi)^3} p^2 \frac{A(\mathbf{p}, \varepsilon)}{\Delta_1(\mathbf{p}, \varepsilon)} \tag{8.75}$$

which gives the same conductivity as (8.13) when put into (8.64). The term $P(\varepsilon - i\delta, \varepsilon + i\delta)$ leads to the important contribution as $T \to 0$ and $n_i \to 0$.

The other term $P(\varepsilon + i\delta, \varepsilon + i\delta)$ should not be neglected. It is complex but not very singular in the limit where $n_i \to 0$. As is evident from the definition (8.61) the singular parts should arise from the Green's function product:

$$G(\mathbf{p}, \varepsilon + i\delta)^2 = G_{\text{ret}}(\mathbf{p}, \varepsilon)^2 = \frac{1}{(\omega - i\Delta)^2} = \frac{\omega^2 - \Delta^2 - 2i\omega\Delta}{(\omega^2 + \Delta^2)^2} \tag{8.76}$$

$$= \frac{1}{\omega^2 + \Delta^2} - \frac{2\Delta^2}{(\omega^2 + \Delta^2)^2} - \frac{2i\omega\Delta}{(\omega^2 + \Delta^2)^2} \tag{8.77}$$

$$= \frac{A}{2\Delta} - \frac{1}{2}A^2 - i\,\text{Re}[G]A \tag{8.78}$$

As $n_i \to 0$ then $A^2 \to A/\Delta$ and the real part of this expression vanishes. The imaginary part becomes $\omega\delta(\omega)$, which is also zero. The singular parts of this expression vanish as $n_i \to 0$. The vertex corrections to $P(\varepsilon + i\delta, \varepsilon + i\delta)$ are not of order unity but of order $O(n_i)$. In this case, the vertex corrections may actually be a series of terms which are successively smaller, so that the vertex corrections may be obtained by just evaluating the first few. The situation is quite different than for $P(\varepsilon - i\delta, \varepsilon + i\delta)$, where one has to solve the vertex equation and sum all the ladder diagrams.

One can evaluate $P(\varepsilon + i\delta, \varepsilon + i\delta)$ by solving a vertex equation similar to (8.69). Define the scalar vertex γ' by $\mathbf{\Gamma}^{(L)}(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta) = \mathbf{p}\gamma'(p, \varepsilon)$, and then

$$\gamma'(p, \varepsilon) = 1 + n_i \int \frac{d^3p'}{(2\pi)^3} \frac{\mathbf{p}\cdot\mathbf{p}'}{p^2} T_{\mathbf{pp}'}(\varepsilon)^2 \gamma'(p', \varepsilon) G_{\text{ret}}(\mathbf{p}', \varepsilon)^2 \tag{8.79}$$

The vertex function $\gamma'(p, \varepsilon)$ is complex, as is the vertex $T_{\mathbf{pp}'}(\varepsilon)^2$, and the product $G_{\text{ret}}(\mathbf{p}', \varepsilon)^2$. There are two coupled equation for the real and imaginary parts. It may be obtained from the *Ward identity* (1950),

$$\mathbf{\Gamma}(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta) = \mathbf{p} + m\nabla_{\mathbf{p}}\Sigma(\mathbf{p}, \varepsilon + i\delta) \tag{8.80}$$

which is an exact identity between the exact vertex function $\mathbf{\Gamma}(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta)$ and the exact retarded self-energy $\Sigma(\mathbf{p}, \varepsilon + i\delta)$ from impurity scattering. This self-energy is not the one derived in Sec. 4.1, since that was only for scattering from a single impurity and only had unperturbed Green's functions $\mathscr{G}^{(0)}$ as the internal lines in the scattering equation. The exact self-energy is found from scattering from all numbers of impurities and using exact Green's functions as internal lines of diagram—although this procedure must be done carefully in order not to count the same contribution twice. The Ward identity is very convenient, since it

permits the vertex function to be obtained from the self-energy function by a simple operation. It will be proved in the next section.

The present derivation has an approximate vertex function which contains only the ladder diagrams. In that case the Ward identity is still valid as long as the self-energy is found using only diagrams where an electron scatters from one impurity at a time.

The Green's function derivation of this result is certainly harder than the transport equation of Sec. 8.1.1. The one advantage of the Green's function method is that there is no need to make the approximation that $n_i \to 0$. The $n_i \to 0$ limit is made implicitly in solving the Boltzmann equation when it is assumed the particles are plane waves except for occasional scattering events from isolated impurities.

There are several lessons to be learned from this summation of ladder diagrams. The ladder diagrams, although they appear in a higher order of perturbation theory, do not lead to terms in the final answer which are smaller than lower-order terms. A term is not necessarily small if it occurs in a higher order of perturbation theory. It is this feature of vertex corrections which means they can never be dismissed without an investigation. They may be small, but one should always check. It would be nice to have some rules of thumb which would establish whether vertex corrections are important. The way to tell is to examine the scattering process which causes the vertex functions. If the two particle states, on each side of the bubble, can scatter quasielastically so that their relative energy changes little, then vertex corrections are large. The vertex correction is basically a potential divided by an energy denominator $V/\Delta E$, where ΔE is the change in energy of the two particles. If it is small, the small denominator will compensate for the small potential V, so that vertex corrections become sizable. Repeated scatterings, as in a series of ladder diagrams, just cause additional powers of the factor $(V/\Delta E)^n$. Vertex corrections are large when the scattering by the potential causes only a small change in the relative energy of the two particles.

8.1.3. Ward Identities

The evaluation of a two-particle correlation function, such as the Kubo formula for the conductivity, often requires an evaluation of a vertex function. The Ward (1950) identity is an exact relationship between the vertex and self-energy functions in the problem. As an example, two types of Ward identities permit the evaluation of the scalar vertex function $\Gamma(\mathbf{p}, ip)$ or the vector vertex function $\mathbf{\Gamma}(\mathbf{p}, ip)$ which satisfy the equations

$$\Gamma(\mathbf{p}, ip) = 1 + \int \frac{d^3p'}{(2\pi)^3} \Gamma(\mathbf{p}', ip)\mathscr{G}(\mathbf{p}', ip)^2 W_{\mathbf{pp}'}(ip, ip) \tag{8.81}$$

$$\mathbf{\Gamma}(\mathbf{p}, ip) = \mathbf{p} + \int \frac{d^3p'}{(2\pi)^3} \mathbf{\Gamma}(\mathbf{p}', ip)\mathscr{G}(\mathbf{p}', ip)^2 W_{\mathbf{pp}'}(ip, ip) \tag{8.82}$$

The Ward identity states that these two functions are given by

$$\Gamma(\mathbf{p}, ip) = 1 - \left[\frac{\partial \Sigma(\mathbf{p}, z)}{\partial z}\right]_{z=ip} \tag{8.83}$$

$$\mathbf{\Gamma}(\mathbf{p}, ip) = \mathbf{p} + m\nabla_p \Sigma(\mathbf{p}, ip) \tag{8.84}$$

An evaluation of the self-energy function $\Sigma(\mathbf{p}, ip)$ permits an easy evaluation of these two vertex functions. These relationships are proved below.

An important point regarding the Ward identities is that they are not useful for evaluating all vertex functions. An example is provided in the last section, where the Ward identities were useful for finding $P(\varepsilon + i\delta, \varepsilon + i\delta)$ but not $P(\varepsilon - i\delta, \varepsilon + i\delta)$. The Ward identities cannot be applied blindly; they must be used only when appropriate. These circumstances are delineated after the identities are proved.

The Ward identities for impurity scattering were derived by Langer (1961). The similar theorems for the electron–phonon interaction were derived by Engelsberg and Schrieffer (1963). We shall prove the result for the ladder diagrams obtained by scattering from a single impurity. In this case, the self-energy diagram is that for scattering from a single impurity:

$$\Sigma(\mathbf{p}, ip) = n_i \left\{ V_{\mathbf{pp}} + \int \frac{d^3p'}{(2\pi)^3} V_{\mathbf{pp}'} V_{\mathbf{p}'\mathbf{p}} \mathscr{G}(\mathbf{p}', ip) + \int \frac{d^3p' d^3p''}{(2\pi)^6} V_{\mathbf{pp}'} V_{\mathbf{p}'\mathbf{p}''} V_{\mathbf{p}''\mathbf{p}'} \mathscr{G}(\mathbf{p}', ip)\mathscr{G}(\mathbf{p}'', ip) + \cdots \right\} \tag{8.85}$$

An important condition is that the Green's functions in this self-energy diagram are those calculated with the self-energy $\mathscr{G} = 1/[ip - \xi_p - \Sigma]$. Equation (8.85) is a self-consistent equation for the self-energy Σ, since it depends functionally on itself. Unfortunately, the Ward identities do not let one avoid solving an integral equation. Instead, one integral equation is exchanged for another. In this sense the Ward identities are not very useful in practice.

Rather than prove the two separate identities (8.83) and (8.84), a general theorem is proved for which these are two limiting cases. The general theorem is obtained by subtracting the expressions (8.85) for $\Sigma(\mathbf{p}, ip) \equiv \Sigma(p)$ by the same result for $\Sigma(\mathbf{p} + \mathbf{q}, ip + i\omega) \equiv \Sigma(p + q)$:

$$\begin{aligned}\Sigma(p+q) - \Sigma(p) = {} & n_i \int \frac{d^3p_1}{(2\pi)^3} V_{\mathbf{pp}_1} V_{\mathbf{p}_1\mathbf{p}} [\mathscr{G}(p_1 + q) - \mathscr{G}(p_1)] \\ & + n_i \int \frac{d^3p_1 d^3p_2}{(2\pi)^6} V_{\mathbf{pp}_1} V_{\mathbf{p}_1\mathbf{p}_2} V_{\mathbf{p}_2\mathbf{p}} [\mathscr{G}(p_1 + q)\mathscr{G}(p_2 + q) - \mathscr{G}(p_1)\mathscr{G}(p_2)] \\ & + n_i \int \frac{d^3p_1 d^3p_2 d^3p_3}{(2\pi)^9} V_{\mathbf{pp}_1} V_{\mathbf{p}_1\mathbf{p}_2} V_{\mathbf{p}_2\mathbf{p}_3} V_{\mathbf{p}_3\mathbf{p}} \\ & \times [\mathscr{G}(p_1 + q)\mathscr{G}(p_2 + q)\mathscr{G}(p_3 + q) - \mathscr{G}(p_1)\mathscr{G}(p_2)\mathscr{G}(p_3)] + \cdots\end{aligned} \tag{8.86}$$

By purely algebraic manipulations, this series can be shown to be identical to

$$\begin{aligned}\Sigma(p+q) - \Sigma(p) &= n_i \int \frac{d^3p_1}{(2\pi)^3} T_{\mathbf{pp}_1}(ip) T_{\mathbf{p}_1+\mathbf{q},\mathbf{p}+\mathbf{q}}(ip + i\omega) [\mathscr{G}(p_1 + q)\mathscr{G}(p_1)] \\ &\quad \times [\Sigma(p_1 + q) - \Sigma(p_1) + \xi_{\mathbf{p}_1+\mathbf{q}} - \xi_{\mathbf{p}_1} - i\omega] \qquad (8.87) \\ &= n_i \int \frac{d^3p_1}{(2\pi)^3} T_{\mathbf{pp}_1}(ip) T_{\mathbf{p}_1+\mathbf{q},\mathbf{p}+\mathbf{q}}(ip + i\omega) [\mathscr{G}(p_1 + q) - \mathscr{G}(p_1)] \qquad (8.88)\end{aligned}$$

This rather startling result may be demonstrated term by term. The first nonvanishing term has the integrand

$$\begin{aligned}V_{\mathbf{pp}_1} V_{\mathbf{p}_1\mathbf{p}} [\mathscr{G}(p_1 + q) - \mathscr{G}(p_1)] &= V_{\mathbf{pp}_1} V_{\mathbf{p}_1\mathbf{p}} \mathscr{G}(p_1 + q)\mathscr{G}(p_1) \\ &\quad \times [\Sigma(p_1 + q) - \Sigma(p_1) + \xi_{\mathbf{p}_1+\mathbf{q}} - \xi_{\mathbf{p}_1} - i\omega]\end{aligned} \tag{8.89}$$

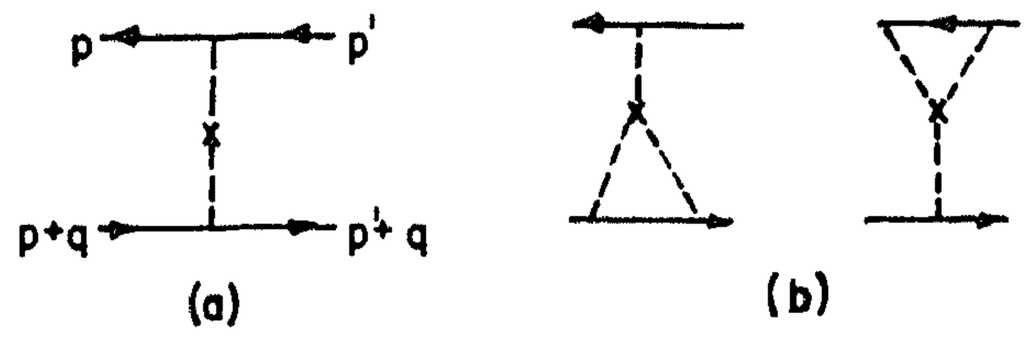


FIGURE 8.5

which is just the vertex diagram shown in Fig. 8.5(a). The next term in the series (8.86) has the integrand

$$V_{\mathbf{p}\mathbf{p}_1} V_{\mathbf{p}_1\mathbf{p}_2} V_{\mathbf{p}_2\mathbf{p}}[\mathscr{G}(p_1+q)\mathscr{G}(p_2+q) - \mathscr{G}(p_1)\mathscr{G}(p_2)] \tag{8.90}$$

The Green's function factors in brackets may be rearranged into

$$\begin{aligned}\mathscr{G}(p_1+q)\mathscr{G}(p_2+q) - \mathscr{G}(p_1)\mathscr{G}(p_2) &= \mathscr{G}(p_1+q)[\mathscr{G}p_2+q) - \mathscr{G}(p_2)] \\ &\quad - \mathscr{G}(p_2)[\mathscr{G}(p_1+q) - \mathscr{G}(p_1)]\end{aligned} \tag{8.91}$$

The first bracket on the right is the first diagram in Fig. 8.5(b). The factor $V_{\mathbf{p}\mathbf{p}_1} V_{\mathbf{p}_1\mathbf{p}_2}\mathscr{G}(p_1+q)$ is just the multiple scattering from the impurity by the electron on the lower line. Similarly, the second term in (8.91) corresponds to the second diagram in Fig. 8.5(b), where the factor $V_{\mathbf{p}_1\mathbf{p}_2} V_{\mathbf{p}_2\mathbf{p}}\mathscr{G}(p_2)$ is the multiple scattering of the top electron line. These terms are just the first in the series which generates the T-matrices. The further terms in (8.86) provide the remaining terms. In this manner, one can establish the validity of (8.88).

Equation (8.87) shows the quantity

$$\Lambda(p, p+q) = \Sigma(p+q) - \Sigma(p) + \xi_{\mathbf{p}+\mathbf{q}} - \xi_{\mathbf{p}} - i\omega \tag{8.92}$$

obeys the vertex equation

$$\begin{aligned}\Lambda(p, p+q) &= \xi_{\mathbf{p}+\mathbf{q}} - \xi_{\mathbf{p}} - i\omega \\ &\quad + n_i \int \frac{d^3p_1}{(2\pi)^3}\mathscr{G}(p_1)\mathscr{G}(p_1+q)T_{\mathbf{p}\mathbf{p}_1}T_{\mathbf{p}_1+\mathbf{q},\mathbf{p}+\mathbf{q}}\Lambda(p_1, p_1+q)\end{aligned} \tag{8.93}$$

The two equations (8.92) and (8.93) provide the most general type of Ward identity. They are useful, since any equation which can be cast into the form of (8.93) has the solution (8.92). Langer (1961) and Engelsberg and Schrieffer (1963) show that this equation is related to the equation of continuity $\nabla \cdot \mathbf{j} + \dot{\rho} = 0$.

The first Ward identity (8.83) is obtained by taking the limit $\mathbf{q} = 0$ and then dividing Eq. (8.93) by $-i\omega$ with the result

$$\begin{aligned}\frac{\Lambda(\mathbf{p}, ip, ip+i\omega)}{-i\omega} &= 1 + n_i \int \frac{d^3p_1}{(2\pi)^3}\mathscr{G}(\mathbf{p}_1, ip)\mathscr{G}(\mathbf{p}_1, ip+i\omega) \\ &\quad \times T_{\mathbf{p}\mathbf{p}_1}(ip)T_{\mathbf{p}_1\mathbf{p}}(ip+i\omega)\frac{\Lambda(\mathbf{p}_1, ip, ip+i\omega)}{-i\omega}\end{aligned} \tag{8.94}$$

The quantity $\Gamma(\mathbf{p}, ip)$ in (8.81) obeys the same equation as $\Lambda/(-i\omega)$ in the limit $i\omega \to 0$ so they are equal. From (8.92) one has the solution

$$\Gamma(\mathbf{p}, ip) = \lim_{i\omega\to 0}\left[\frac{\Lambda(\mathbf{p}, ip, ip+i\omega)}{-i\omega}\right] = 1 - \left[\frac{\partial\Sigma(\mathbf{p}, Z)}{\partial Z}\right]_{Z=ip} \tag{8.95}$$

which proves the Ward identity (8.83).

The other Ward identity is found as the limit $i\omega = 0$, followed by letting $\mathbf{q} \to 0$. The latter limit is taken slowly, so that one can retain terms proportional to $\mathbf{q}$. In this limit Eqns. (8.92) and (8.93) become

$$\lim_{\mathbf{q}\to 0} \Lambda = \frac{\mathbf{q}\cdot\mathbf{p}}{m} + \mathbf{q}\cdot\nabla_{\mathbf{p}}\Sigma(\mathbf{p}, ip) \tag{8.96}$$

$$= \frac{\mathbf{q}\cdot\mathbf{p}}{m} + n_i \int \frac{d^3p_1}{(2\pi)^3} \mathscr{G}(\mathbf{p}_1, ip)\mathscr{G}(\mathbf{p}_1 + \mathbf{q}, ip) T_{\mathbf{p}\mathbf{p}_1} T_{\mathbf{p}_1+\mathbf{q},\mathbf{p}+\mathbf{q}} \Lambda(p_1, p_1 + q) \tag{8.97}$$

The vertex function Λ is proportional to $\mathbf{q}$, so define the vector vertex function by the limit ($i\omega = 0$):

$$\lim_{\mathbf{q}\to 0} \Lambda = \frac{1}{m}\mathbf{q}\cdot\mathbf{\Gamma}(\mathbf{p}, ip) \tag{8.98}$$

Then the preceding two equations can be expressed in terms of this vector vertex function:

$$\mathbf{\Gamma}(\mathbf{p}, ip) = \mathbf{p} + m\nabla_{\mathbf{p}}\Sigma(\mathbf{p}, ip) \tag{8.99}$$

$$\mathbf{\Gamma}(\mathbf{p}, ip) = \mathbf{p} + n_i \int \frac{d^3p_1}{(2\pi)^3} \mathscr{G}(\mathbf{p}_1, ip)^2 |T_{\mathbf{p}\mathbf{p}_1}(ip)|^2 \mathbf{\Gamma}(\mathbf{p}_1, ip) \tag{8.100}$$

This equation is the same as (8.82), which proves the other Ward identity (8.84). Both are now understood to be limiting cases of the general result (8.93). The Ward identities are useful anytime one can cast the vertex equation into the form (8.93).

The factor $[1 - \partial\Sigma/\partial\Sigma]$ is recognized as the inverse of the renormalization Z defined earlier and discussed, for example, in Sec. 5.8.1. This quantity is sometimes called the effective charge. Similarly, the vector vertex is one of the factors which give the effective mass of the particle. The Ward identities relate the vertex corrections to a change in the effective charge and mass of the particle, which is why they are related to the equation of continuity.

8.2. MOBILITY OF FRÖHLICH POLARONS

The Fröhlich Hamiltonian between electrons and Einstein phonons ($\omega_0 = \omega_{LO}$) is

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \omega_0 \sum_{\mathbf{q}} a^\dagger_{\mathbf{q}} a_{\mathbf{q}} + \frac{M_0}{\sqrt{\nu}} \sum_{\mathbf{p}\mathbf{q}} \frac{1}{|\mathbf{q}|} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma}(a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}}) \tag{8.101}$$

$$M_0^2 = \frac{4\pi\alpha\hbar(\hbar\omega_0)^{3/2}}{\sqrt{2m_B}}, \qquad \varepsilon_{\mathbf{p}} = \frac{p^2}{2m_B}$$

$$\alpha = \frac{e^2}{\hbar}\left(\frac{m_B}{2\hbar\omega_0}\right)^{1/2}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right) \tag{8.102}$$

It was discussed in Sec. 7.1, where several important quantities were derived, such as the effective mass m^* and the ground state energy E_0.

The effective mass m^* of a particle can be measured by cyclotron resonance. Such experiments have been done for polarons (Hodby, 1972). The effective mass m^* is a function of the band mass m_B and the polar coupling constant α. A separate measurement of the two dielectric functions ε_0 and ε_∞, as well as ω_{LO}, permits a determination of the band mass m_B

and α from m^*. This analysis takes a theory of the polaron mass $m^*(m_B, \alpha)$, which was provided in Sec. 7.1. Another way to check the theory is to measure the mobility of electrons in insulators. The mobility also depends upon m_B and α.

A typical experimental result for the mobility is shown in Fig. 8.6. It shows the Hall mobility of CdTe measured by Segall *et al.* (1963). The steep rise around 200 K is due to optical mode scattering. At lower temperatures the mobility saturates because of the scattering from impurities in the crystals. Impurity scattering varies from sample to sample, depending on the concentration and type of impurity.

The average value of the current operator is the particle density n_0 times the charge e times the average velocity $\langle \mathbf{v} \rangle$. The average velocity $\langle \mathbf{v} \rangle$ is proportional to the applied electric field $\mathbf{F}$, and the constant of proportionality is the mobility, where $e > 0$ so the electron charge is $-e$:

$$\langle \mathbf{v} \rangle = \mu \mathbf{F} \tag{8.103}$$

$$\mathbf{J} = -e n_0 \mu \mathbf{F} \tag{8.104}$$

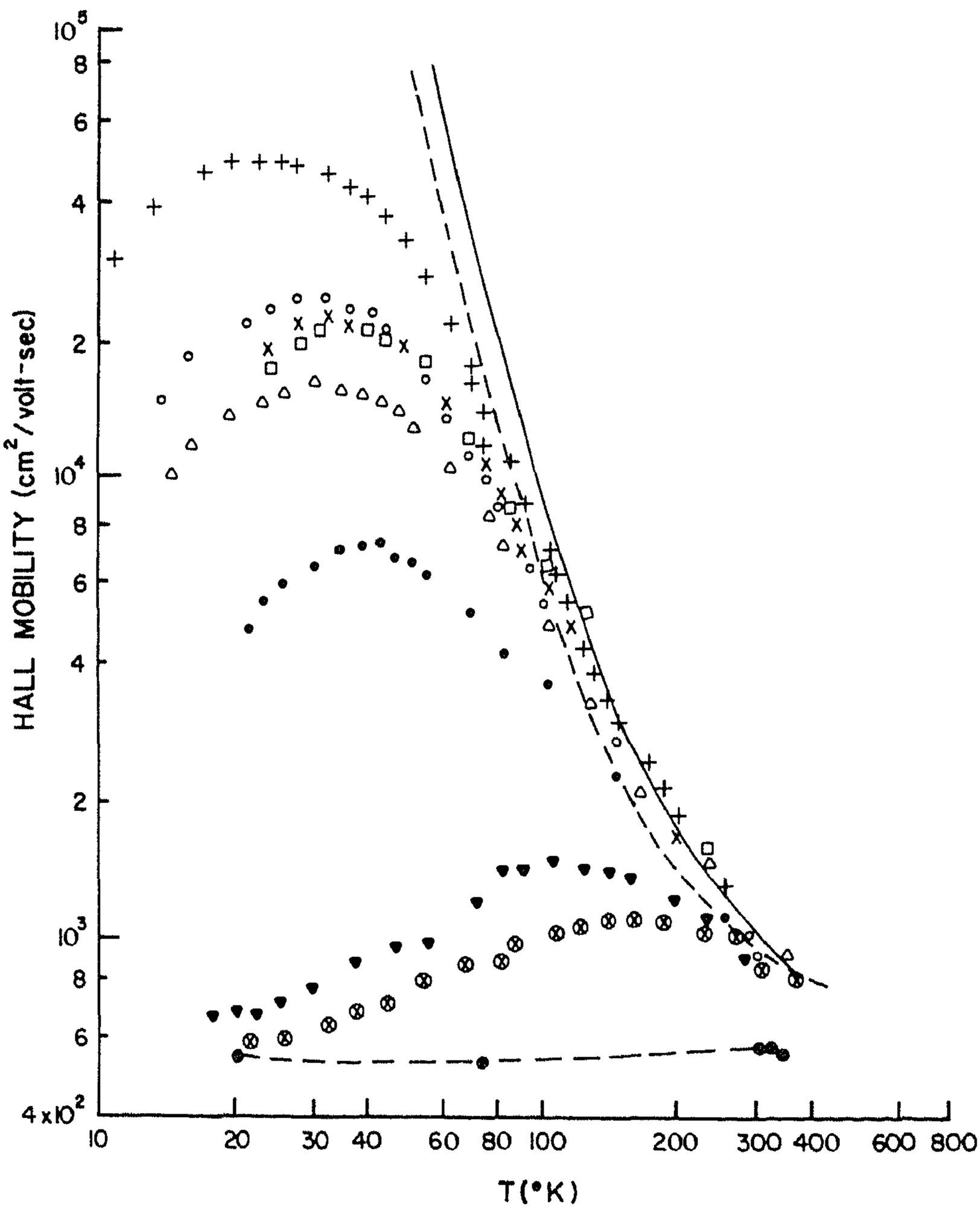


FIGURE 8.6 The temperature dependence of the electron mobility of several samples of n-type CdTe. Different samples have different kinds and concentrations of defects or impurities. The solid line is the theoretical mobility from optical mode scattering, including the temperature dependence of the static dielectric constant. The dashed line neglects this change. *Source:* Segall *et al.* (1963) (used with permission).

The mobility μ is the average velocity of each electron per unit applied electric field. Of course, it is strictly defined in the limit of vanishing electric field. Since the electrical conductivity $\sigma = n_0 e^2 \tau_t / m = -e\mu n_0$ is the ratio of the current $\mathbf{J}$ to the field $\mathbf{F}$, then $\mu = -e\tau_t/m$.

The theories of the electron mobility in insulating materials, such as alkali halides and II–VI semiconductors, treat it as a property of a single electron. The electron lifetime is calculated for the scattering from impurities and by acoustical and optical phonons. The electron–electron interactions can be ignored in the limit where the concentration of electrons is very low. Spin is also unimportant, and the spin index is omitted in this section.

There are just as many different ways to calculate the polaron mobility as there are to calculate the effective mass m^* or ground state energy E_0. Each theoretical technique was applied to the mobility as well as to the other quantities. As summarized by Langreth (1967), these various methods usually agree in the limit of weak coupling ($\alpha \ll 1$) and low temperature $k_B T \ll \hbar\omega_0$. In this case

$$\lim_{T\to 0} \mu = -\frac{e\tau_0}{m_B} \equiv \mu_0 \tag{8.105}$$

$$\frac{1}{\tau_0} = 2\alpha N_0 \omega_0, \qquad N_0 = \frac{1}{e^{\beta\omega_0} - 1} \tag{8.106}$$

The lifetime τ_0 is the result obtained in (7.46) as $1/\tau_0 = -2\,\mathrm{Im}[\Sigma^{(1)}(p, \varepsilon)]$ in the limit where $p \to 0$ and $\varepsilon \to 0$, where $\Sigma^{(1)}$ is the one-phonon self-energy. This limit is appropriate, since at very low temperatures the electrons are in states within $k_B T \ll \hbar\omega_0$ of the bottom of the band. These low-energy particles cannot emit phonons, since this event is prevented by energy conservation. They can only absorb them, and the rate of absorption is proportional to the thermal average density of phonons N_0. The factor N_0 makes the mobility increase exponentially with decreasing temperature, since the electron scattering becomes less likely as the number density of phonons declines. The exponential increase in the mobility is evident in the experimental data of Fig. 8.6. The behavior of large polarons is opposite to that of small polarons, whose mobility increases with increasing temperature.

One feature of the mobility formula (8.105) is that it is proportional to the inverse of α. Our starting point for the theoretical calculation is again the Kubo formula, which will be evaluated for using electron–phonon interactions. The expansion of the S matrix for this potential will generate a series in the parameter α. To obtain a leading term in the inverse power of α requires the summation of a subset of diagrams. The situation is similar to the mobility from impurity scattering, where diagrams were summed to get the conductivity inversely proportional to the impurity concentration n_i. There are important differences between the lifetime from opical phonon scattering and that of impurity scattering. This conclusion is evident from the result presented in (8.105). Here the relaxation time is not calculated with the factor of $(1 - \cos\theta')$ in the angular average. The polaron mobility is calculated in a different way than the scattering from impurities. Actually it is calculated in the same way, but a different result is obtained. This difference arises from the inelastic nature of the polaron scattering as first shown by Howarth and Sondheimer (1953).

Langreth and Kadanoff (1964) showed that the polaron mobility μ is a power series in α, with the leading term in (8.105) of order $1/\alpha$:

$$\mu = \frac{a_{-1}}{\alpha} + a_0\alpha^0 + a_1\alpha^1 + \cdots \tag{8.107}$$

They tried to calculate the coefficient a_{-1}, $a_0$in this series in the limit of low temperature. This objective is simple but has slippery aspects (Mahan, 1966). It is a subtle procedure to examine each term in the S-matrix expansion and to determine its leading term in α. The situation is similar to that for impurity scattering. Each term in the series (8.107) is obtained by summing subsets of diagrams.

The first term of Langreth and Kadanoff (a_{-1}) is quite simple, and it is probably worthwhile to state it in advance. Define μ_0 as the result (8.105) for the limits $\alpha \to 0$, $T \to 0$. They found

$$\frac{\mu}{\mu_0} = 1 - \frac{\alpha}{6} + O(\alpha^2) \tag{8.108}$$

They observed that μ_0/μ is precisely the expansion given by the equations

$$\mu = -\frac{e\tau}{m^*} \tag{8.109}$$

$$\frac{\tau}{\tau_0} = 1 + O(\alpha^2) \tag{8.110}$$

$$\frac{m^*}{m_B} = 1 + \frac{\alpha}{6} + O(\alpha^2) \tag{8.111}$$

The ballistic formula $\mu = -e\tau/m^*$ supports the quasiparticle picture that the particle acts as if it has an effective mass m^* and lifetime τ. They speculate that the quasiparticle picture would be valid for all values of α and that the inclusion of all terms in α would just reproduce the product series of m^* and τ.

Before the derivation of the electron mobility, it is necessary to derive some single-particle properties. These will be needed in the limit of zero temperature. All terms are dropped of order N_0 compared to unity. Many of these single-particle properties were derived in Sec. 7.1. The first self-energy term, proportional to α, is the one-phonon result in (7.15):

$$\mathrm{Re}[\Sigma^{(1)}(p, \omega)] = -\alpha \frac{\omega_O^{3/2}}{\sqrt{\varepsilon_p}} \sin^{-1}\left(\frac{\varepsilon_p}{\omega_0 - \omega + \varepsilon_p}\right)^{1/2} \tag{8.112}$$

At zero temperature, this self-energy is evaluated at small ε_p and small ω,

$$\mathrm{Re}[\Sigma^{(1)}(p, \omega)] = -\alpha[\omega_0 + \tfrac{1}{2}\omega - \tfrac{1}{3}\varepsilon_p + O(\omega^2, \varepsilon_p^3, \omega\varepsilon_p)] \tag{8.113}$$

This expansion permits a quick derivation of the effective mass m^* and the renormalization coefficient Z. The zero subscript means $p = 0$, $\omega = 0$:

$$Z_0 = \left(1 - \frac{\partial\Sigma}{\partial\omega}\right)_0^{-1} = \frac{1}{1 + \alpha/2} \approx 1 - \frac{\alpha}{2} + O(\alpha^2) \tag{8.114}$$

$$\left(\frac{m_B}{m^*}\right)_0 = Z_0\left(1 + \frac{\partial\Sigma}{\partial\varepsilon_p}\right) = \frac{1 + \alpha/3}{1 + \alpha/2} \approx 1 - \frac{\alpha}{6} + O(\alpha^2) \tag{8.115}$$

Another important quantity is the lifetime τ which is defined as

$$\frac{1}{\tau(p)} = Z(p)\{-2\,\mathrm{Im}[\Sigma(p, E_p)]\} \tag{8.116}$$

where $Z(p)$ is the renormalization coefficient evaluated at $\omega = E_p$, which is the ground state energy. The ground state energy is only needed to order α, which is the simple result $E_p = -\alpha\omega_0 + p^2/2m^* + O(p^4)$.

Equation (8.116) shows that the renormalization coefficient $Z(p)$ enters into the definition of the lifetime. The argument for this is as follows. The spectral function is defined as

$$A(p, \omega) = -2\,\mathrm{Im}[G_{\mathrm{ret}}(p, \omega)] \tag{8.117}$$

$$= \frac{-2\,\mathrm{Im}[\Sigma(p, \omega)]}{\{\omega - \varepsilon_p - \mathrm{Re}[\Sigma(p, \omega)]\}^2 + \{\mathrm{Im}[\Sigma(p, \omega)]\}^2} \tag{8.118}$$

A suitable definition of $\tau(p)$ is obtained by examining this limit more carefully when $\mathrm{Im}(\Sigma)$ is small but not infinitesimal. In the vicinity of the peak $\omega \approx E_p$ of the spectral function the term in the denominator is

$$\omega - \varepsilon_p - \mathrm{Re}[\Sigma(p, \omega)] \approx \omega - \varepsilon_p - \mathrm{Re}[\Sigma(p, E_p)] - (\omega - E_p)\frac{\partial\,\mathrm{Re}\,\Sigma}{\partial\omega}$$

$$\approx (\omega - E_p)\left[1 - \frac{\partial\,\mathrm{Re}\,\Sigma}{\partial\omega}\right] = \frac{(\omega - E_p)}{Z(p)} \tag{8.119}$$

so that the spectral function is approximately

$$A(p, \omega) = \frac{-2\,\mathrm{Im}[\Sigma(p, \omega)]}{(\omega - E_p)^2/Z(p)^2 + [\mathrm{Im}(\Sigma)]^2} \tag{8.120}$$

$$= Z(p)\left(\frac{1/\tau(p)}{((\omega - E_p)^2 + (1/2\tau)^2}\right) \tag{8.121}$$

The last step used the definition (8.116) of the relaxation time. The relaxation time $\tau(p)$ is treated as a function of p but not ω. This form for the spectral function is used in the Green's function (Problem 6 in Chapter 3),

$$G_{\mathrm{ret}}(p, t) = \Theta(t)\int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} e^{-i\omega t} A(p, \omega) \tag{8.122}$$

$$= -iZ(p)\Theta(t)\exp[-itE_p - t/(2\tau)] \tag{8.123}$$

The relaxation time is defined from the decay of the Green's function. It has the desired form, with the relaxation time $\tau(p)$ determining the decay of the excitation.

There is another way to understand the factor Z in the definition (8.116) of $\tau(p)$. The quantity $-2\,\mathrm{Im}(\Sigma)$ is the rate of decay of a state (p, ω). The factor Z is the fraction of the quasiparticle strength at the value (p, ω). The rest of the quasiparticle strength is usually dispersed throughout the spectrum.

The one-phonon self-energy is evaluated for the quasiparticle lifetime. The imaginary self-energy is evaluated at the quasiparticle energy $\omega = E_0 = -\alpha\omega_0$ and is multiplied by the factor of Z given in (8.115). The imaginary self-energy is calculated from the expression

$$-2\,\mathrm{Im}[\Sigma(p,\omega)] = 2\pi \int \frac{d^3q}{(2\pi)^3}\frac{M_0^2}{q^2} \tag{8.124}$$

$$\times\,[N_0\delta(\omega+\omega_0-\varepsilon_{\mathbf{q}+\mathbf{p}}) + (N_0+1)\delta(\omega-\omega_0-\varepsilon_{\mathbf{q}+\mathbf{p}})]$$

$$= \frac{\alpha(\omega_0)^{3/2}}{\sqrt{\varepsilon_p}}\left[N_0\Theta(\omega+\omega_0)\ln\left|\frac{\sqrt{\omega+\omega_0}+\sqrt{\varepsilon_p}}{\sqrt{\omega+\omega_0}-\sqrt{\varepsilon_p}}\right|\right.$$

$$\left. + (N_0+1)\Theta(\omega-\omega_0)\ln\left|\frac{\sqrt{\omega-\omega_0}+\sqrt{\varepsilon_p}}{\sqrt{\omega-\omega_0}-\sqrt{\varepsilon_p}}\right|\right] \tag{8.125}$$

Expanding for small $\omega < \omega_0$ and small ε_p gives

$$-2\,\mathrm{Im}[\Sigma(0,\omega)] = \frac{2\alpha N_0\omega_0^{3/2}}{\sqrt{\omega_0+\omega}} \approx 2\alpha N_0[\omega_0 - \tfrac{1}{2}\omega + \cdots] \tag{8.126}$$

$$-2\,\mathrm{Im}[\Sigma(0,\omega=-\alpha\omega_0)] = 2\alpha N_0\omega_0\left[1+\frac{\alpha}{2}+O(\alpha^2)\right] \tag{8.127}$$

$$\frac{1}{\tau} = Z[-2\,\mathrm{Im}(\Sigma)] = \frac{1}{\tau_0}\left(1-\frac{\alpha}{2}\right)\left(1+\frac{\alpha}{2}\right)$$

$$= \frac{1}{\tau_0}[1+O(\alpha^2)] \tag{8.128}$$

The first correction terms in α from $Z = 1-\alpha/2$ and $\mathrm{Im}[\Sigma] = 1+\alpha/2$ cancel to order $O(\alpha)$. The one-phonon term provides no correction term to τ of $O(\alpha)$.

The electron mobility is evaluated be the same method which was used for impurity scattering. Starting from the Kubo formula for the electrical conductivity, the first important diagram is the bubble with interacting Green's functions shown in Fig. 8.7(a). The solid lines are total Green's functions $\mathscr{G}(p, ip_m)$ which include the self-energies

$$\pi(i\omega) = -\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \mathbf{j}(\tau)\cdot\mathbf{j}(0)\rangle \tag{8.129}$$

$$\pi^{(0)}(i\omega) = \frac{2e^2}{3m_B^2}\int\frac{d^3p}{(2\pi)^3}p^2\frac{1}{\beta}\sum_{ip_m}\mathscr{G}(\mathbf{p},ip_m)\mathscr{G}(\mathbf{p},ip_m+i\omega_n) \tag{8.130}$$

The solid lines are total Green's functions $\mathscr{G}(\mathbf{p}, ip_m)$ which include the self-energies. The evaluation of this term is identical to that used for the same bubble diagram for impurity

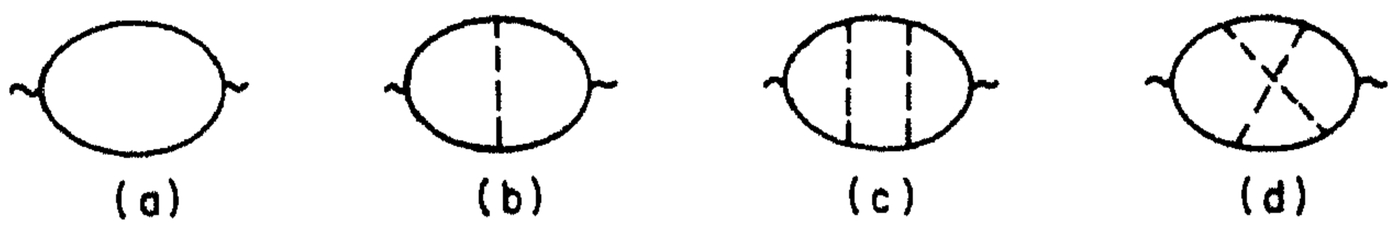


FIGURE 8.7

scattering. The contribution to the Kubo formula contains the two Green's functions as in Eq. (8.44), which are evaluated by the standard series of steps to yield Eq. (8.49):

$$\sigma^{(0)} = \frac{e^2}{3m_B^2}\int\frac{d^3p}{(2\pi)^3}p^2\int\frac{d\varepsilon}{2\pi}A(p,\varepsilon)^2\left[-\frac{dn_F(\varepsilon)}{d\varepsilon}\right] \tag{8.131}$$

In the present application, the self-energies in the Green's functions are evaluated from the Fröhlich Hamiltonian rather than impurity scattering. In a more realistic model of a solid, both self-energy expressions should be included: from phonons and impurity scattering.

An approximate evaluation of this contribution to the conductivity is obtained in the limit where $\alpha \ll 1$. The electron distribution is assumed to be Maxwell–Boltzmann:

$$n_F(\varepsilon) \approx e^{\beta(\mu-\varepsilon)} = \frac{n_0}{2}\left(\frac{2\pi\beta}{m^*}\right)^{3/2} e^{-\beta\varepsilon} \tag{8.132}$$

where $n_F(\varepsilon)$ is the energy distribution for each spin state, while n_0 is the total concentration for both spin states. This term in the conductivity is divided by $-en_0$ to get the corresponding term in the mobility of each electron:

$$\mu^{(0)} = -\frac{e\beta}{6m_B^2}\left(\frac{2\pi\beta}{m^*}\right)^{3/2}\int\frac{d^3p}{(2\pi)^3}p^2\int\frac{d\varepsilon}{2\pi}A(p,\varepsilon)^2 e^{-\beta\varepsilon} \tag{8.133}$$

Using the expression (8.121) for the spectral function, the square of the spectral function is

$$A(p,\varepsilon)^2 = \frac{(Z/\tau)^2}{\{(\varepsilon-E_p)^2+(2\tau)^{-2}\}^2} \approx A\pi\tau Z^2\delta(\varepsilon-E_p) \tag{8.134}$$

The electron lifetime τ is inversely proportional to α, so that the mobility $\mu^{(0)}$ is inversely proportional to α. The preceding integral is evaluated in the limit where the temperature $T \to 0$. The integral is easy if it is assumed at low T that $Z(p)$ and $\tau(p)$ are evaluated at $p = 0$

$$\mu^{(0)} = -e\tau Z^2\frac{m^*}{m_B^2} = -\frac{e\tau_0}{m_B}(1-\tfrac{5}{6}\alpha+O(\alpha^2)) \tag{8.135}$$

where $Z(0) = 1-\alpha/2$ and $m^*/m_B = 1+\alpha/6$. The result (8.135) is the contribution from the simple bubble diagram of Fig. 8.8(a). This result does not resemble (8.108). The differences disappear when higher-order diagrams are considered, such as those in Fig. 8.7.

Other contributions to μ can be derived from the other diagrams, which are the vertex corrections. In the limits $T \to 0$ and $\alpha \ll 1$ the vertex corrections do not contribute to the mobility a term which goes as $O(1/\alpha)$. The simple bubble result is the final answer at low temperature and weak coupling. This conclusion is quite different from the situation for impurity scattering. There a series of ladder diagrams was summed in order to derive the final answer, and each ladder diagram gave a term which was the same inverse power of the coupling constant as the impurity concentration n_i. The vertex diagrams for optical phonon scattering are less important because of the inelastic nature of the phonon scattering. Mahan (1966) showed that the two-phonon ladder diagrams in Fig. 8.7(c) provide the largest vertex corrections, which are of $O(T/\omega_0)$.

8.3. ELECTRON–PHONON RELAXATION TIMES

This section is concerned with the calculation of electron relaxation times in semiconductors from the scattering by acoustical phonons. The prior sections have introduced two different relaxation times for the electron in a solid. One of these is the average time between scattering events and is denoted as $\tau(k)$. The mean-free-path (mfp) for this relaxation time is denoted as $l(k) = v_k\tau(k)$. The energy bands in the solid are assumed to be isotropic, so that the relaxation time and mfp depend only upon the magnitude of the wave vector. The other relaxation time τ_t enters into the electrical conductivity ($\sigma = n_0e^2\tau_t/m^*$). The equivalent quantity, before averaging over wave vector, is called the momentum relaxation time and is denoted with the subscript "t" since it is used in transport of current. The equivalent mfp is $l_t(k) = v_k\tau_t(k)$.

To further confuse the topic, the transport lifetimes τ_t for scattering from impurities had a factor of $(1 - \cos\theta')$ in the integrand of the scattering integral. The scattering from optical phonons did not have such a factor. The difference is that at low temperature, impurity scattering is elastic, while optical phonon is highly inelastic. Vertex corrections are relatively unimportant for inelastic scattering.

In writing the Boltzmann equation, the last term in the time development is the rate of change of the distribution function from collisions. The discussion of the lifetime from impurity scattering in Sec. 8.1.1 started with this collision term. Here the similar collision term is presented from the scattering due to the electron–phonon interaction:

$$\begin{aligned}\left(\frac{\partial f(\mathbf{p})}{\partial t}\right)_{ep} &= \frac{2\pi}{\hbar}\int\frac{d^3q}{(2\pi)^3}|M_{\mathbf{q}}|^2\{f(\mathbf{p})[1-f(\mathbf{p}+\mathbf{q})]\\ &\times[(N_{\mathbf{q}}+1)\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}+\mathbf{q}}-\hbar\omega_{\mathbf{q}})+N_{\mathbf{q}}\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}+\mathbf{q}}+\hbar\omega_{\mathbf{q}})]\\ &-f(\mathbf{p}+\mathbf{q})[1-f(\mathbf{p})][(N_{\mathbf{q}}+1)\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}+\mathbf{q}}+\hbar\omega_{\mathbf{q}})+N_{\mathbf{q}}\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}+\mathbf{q}}-\hbar\omega_{\mathbf{q}})]\}\end{aligned} \tag{8.136}$$

The integrand has four terms. The first two correspond to an electron initially in state $\mathbf{p}$ scattering to $\mathbf{p}+\mathbf{q}$ by either the emission $(N_{\mathbf{q}}+1)$ or absorption $(N_{\mathbf{q}})$ of a phonon. This term is multiplied by the occupation factors $f(\mathbf{p})[1-f(\mathbf{p}+\mathbf{q})]$ which ensures that the initial state $\mathbf{p}$ is occupied and the final state is empty. The other two terms correspond to processes whereby electrons initially in $\mathbf{p}+\mathbf{q}$ scatter back into the state $\mathbf{p}$ by either phonon emission or absorption. The back scattering has the occupation factors of $f(\mathbf{p}+\mathbf{q})[1-f(\mathbf{p})]$. The above expression vanishes when the system is in thermal equilibrium and $f(\mathbf{p}) = n_F(\xi_{\mathbf{p}})$, $N_{\mathbf{q}} = n_B(\omega_{\mathbf{q}})$.

An immediate simplification of the above formula is attained by grouping together the terms with the same delta function for energy conservation

$$\begin{aligned}\left(\frac{\partial f(\mathbf{p})}{\partial t}\right)_{ep} &= \frac{2\pi}{\hbar}\int\frac{d^3q}{(2\pi)^3}|M_{\mathbf{q}}|^2\\ &\times[\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}+\mathbf{q}}-\hbar\omega_{\mathbf{q}})\{f(\mathbf{p})[N_{\mathbf{q}}+1-f(\mathbf{p}+\mathbf{q})]-N_{\mathbf{q}}f(\mathbf{p}+\mathbf{q})\}\\ &+\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}+\mathbf{q}}+\hbar\omega_{\mathbf{q}})]\{f(\mathbf{p})[N_{\mathbf{q}}+f(\mathbf{p}+\mathbf{q})]-f(\mathbf{p}+\mathbf{q})[N_{\mathbf{q}}+1]\}]\end{aligned} \tag{8.137}$$

The above expression is quite general.

The terms which multiply $f(\mathbf{p})$are those which define the relaxation time $1/\tau_p(\mathbf{p})$.

$$\frac{1}{\tau_p(k)} = \frac{2\pi}{\hbar}\int \frac{d^3q}{(2\pi)^3}|M_q|^2\{\delta(\varepsilon_k - \varepsilon_{\mathbf{k+q}} - \hbar\omega_{\mathbf{q}})[N_q + 1 - n_F(\varepsilon_{\mathbf{k+q}})]$$
$$+ \delta(\varepsilon_k - \varepsilon_{\mathbf{k+q}} + \hbar\omega_{\mathbf{q}})[N_q + n_F(\varepsilon_{\mathbf{k+q}})]\} \tag{8.138}$$

$$\frac{1}{\tau_{pt}(k)} = \frac{2\pi}{\hbar}\int \frac{d^3q}{(2\pi)^3}|M_q|^2\left(-\frac{\mathbf{q}\cdot\mathbf{k}}{k^2}\right)$$
$$\times \{\delta(\varepsilon_k - \varepsilon_{\mathbf{k+q}} - \hbar\omega_{\mathbf{q}})[N_q + 1 - n_F(\varepsilon_{\mathbf{k+q}})]$$
$$+ \delta(\varepsilon_k - \varepsilon_{\mathbf{k+q}} + \hbar\omega_{\mathbf{q}})[N_q + n_F(\varepsilon_{\mathbf{k+q}})]\} \tag{8.139}$$

$$|M_q|^2 = D^2\frac{\hbar q^2}{2\rho\omega_q} \tag{8.140}$$

$$N_q = n_B(\omega_q) \tag{8.141}$$

The subscript "p" denotes scattering by phonons. In Eq. (8.138) the relaxation time from phonon scattering is all of the terms which multiply $f(\mathbf{p})$ in (8.137). It is identical to the expression obtained from the imaginary part of the electron self-energy, as calculated in the one-phonon approximation. The occupation factors $f(\mathbf{p}+\mathbf{q})$, $N_{\mathbf{q}}$ are represented by their values in thermal equilibrium. The matrix element is from deformation potential interactions.

The transport lifetime τ_{pt} from the electron scattering by phonons is defined in (8.140). It has an additional factor of $(-\mathbf{q}\cdot\mathbf{k}/k^2)$ in the integrand. If the scattering were elastic, then this factor equals $(1-\cos\theta')$. The difference between the impurity scattering times τ_i, and τ_{it}, is the factor of $[1-\cos(\theta)]$ in the integrand of $1/\tau_{it}$, where θ is the scattering angle $\hat{k}\cdot\hat{p} = \cos(\theta)$.

$$\frac{1}{\tau_i(k)} = \frac{2\pi n_i}{\hbar}\int \frac{d^3p}{(2\pi)^3}|T_{\mathbf{kp}}|^2\delta(\varepsilon_k - \varepsilon_p) \tag{8.142}$$

$$\frac{1}{\tau_{it}(k)} = \frac{2\pi n_i}{\hbar}\int \frac{d^3p}{(2\pi)^3}|T_{\mathbf{kp}}|^2[1-\hat{k}\cdot\hat{p}]\delta(\varepsilon_k - \varepsilon_p) \tag{8.143}$$

The identical factor is used to calculate the momentum relaxation (τ_{pt}) from scattering by phonons. If $\mathbf{p} = \mathbf{k}+\mathbf{q}$ then the angular factor $[1-\cos(\theta)] \approx -\mathbf{q}\cdot\mathbf{k}/k^2$. The identity is exact if $p = k$. Using this factor of $[1-\cos\theta]$ in the impurity scattering is rigorously correct. Using this factor in the scattering by phonons is not rigorously correct. The difference is that the scattering by phonons is inelastic. Instead, the correct result is found by solving the Boltzmann equation, which is done in Sec. 8.4. However, the above approximation is actually quite good, and therefore is useful. The lifetimes for scattering by phonons are evaluated below.

Besides these two relaxation times, there are several others which are occasionally useful. The most important is the scattering by electron–electron interactions in a metal. The formulas for this case are derived in Chapter 11 in the discussion of Fermi liquid theory. Two others are the relaxation times for temperature, and the relaxation times for energy. These two are discussed below.

The relaxation time for energy determines the rate at which the electron loses or gains energy from the scattering. Since impurity scattering is elastic, the electron does not change its energy. There is no contribution to energy relaxation from impurity scattering. However,

in the scattering by phonons the electron changes its energy by $\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}} = \pm\hbar\omega_{\mathbf{q}}$. One process gains energy $(+\hbar\omega_{\mathbf{q}})$ while the other loses energy $(-\hbar\omega_{\mathbf{q}})$. There is a net gain or loss of energy if one process dominates over the other. For example, if an energetic electron is injected into the solid, it will gradually lose energy until it equilibrates thermally. The process of coming to thermal equilibrium requires that it emit more phonons than it absorbs.

The rate of energy relaxation is calculated by starting from (8.138) and inserting the energy change $\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}$ into the integrand

$$\left(\frac{dE}{dt}\right)_{pe} = -\frac{2\pi}{\hbar}\int\frac{d^3q}{(2\pi)^3}|M_q|^2\hbar\omega_q\{\delta(\varepsilon_k - \varepsilon_{\mathbf{k}+\mathbf{q}} - \hbar\omega_{\mathbf{q}})[N_q + 1 - n_F(\varepsilon_{\mathbf{k}+\mathbf{q}})]$$
$$- \delta(\varepsilon_k - \varepsilon_{\mathbf{k}+\mathbf{q}} + \hbar\omega_{\mathbf{q}})[N_q + n_F(\varepsilon_{\mathbf{k}+\mathbf{q}})]\} \tag{8.144}$$

Converting this expression to a relaxation time (τ_{pe}) and mfp (l_{pe}) is done below.

The evaluation of the relaxation times for the electron–phonon systems is done for metals and semiconductors. In metals, the evaluation assumes the existence of a Fermi surface. The electron–phonon scattering affects those few electrons within about k_BT of the Fermi surface. In semiconductors, the interesting effects are for a single electron in the band, and the reference energy is the band edge. These two cases are different and are treated separately.

8.3.1. Metals

In metals the derivation is simplified by introducing the McMillan function denoted as $\alpha^2F(\omega)$. In this function are collected all of the complicated parts of the phonon dispersion, and the matrix element (see Sec. 7.4)

$$\alpha^2F(E,\omega) = \int\frac{d^3q}{(2\pi)^3}|M_q|^2\delta(\omega - \omega_{\mathbf{q}})\delta(E - \varepsilon_{\mathbf{k}+\mathbf{q}}) \tag{8.145}$$

$$\alpha_t^2F(E,\omega) = \int\frac{d^3q}{(2\pi)^3}|M_q|^2\left(-\frac{\mathbf{q}\cdot\mathbf{k}}{k^2}\right)\delta(\omega - \omega_{\mathbf{q}})\delta(E - \varepsilon_{\mathbf{k}+\mathbf{q}}) \tag{8.146}$$

$$\frac{1}{\tau_{p(t)}(k)} = \frac{2\pi}{\hbar}\int d\omega\{\alpha_{(t)}^2F(\varepsilon_k - \omega,\omega)[n_B(\omega) + n_F(\omega - \varepsilon_k)]$$
$$+ \alpha_{(t)}^2F(\varepsilon_k + \omega,\omega)[n_B(\omega) + n_F(\omega + \varepsilon_k)]\} \tag{8.147}$$

The dimensionless functions $\alpha_{(t)}^2F(E,\omega)$ depends upon the variable E. However, this dependence is similar to the variation in the density of states at the Fermi surface. The variation with E is usually smooth on the energy scale of the Debye energy, which is all that is relevant for interactions with phonons. The important variation is

$$\alpha_{(t)}^2F(\varepsilon_k \pm \omega,\omega) = \alpha_{(t)}^2F(\varepsilon_k,\omega) \pm \omega\frac{\partial}{\partial\varepsilon_k}\alpha_{(t)}^2F(\varepsilon_k,\omega) + \cdots \tag{8.148}$$

The second term is usually smaller than the first by a factor of $\hbar\omega/W$, where $\hbar\omega$ is a phonon energy while W is an electronic bandwidth. The usual approximation is to neglect the dependence upon the factor of E The above expression is rewritten as

$$\frac{1}{\tau_{p(t)}(k)} = \frac{2\pi}{\hbar}\int_0^{\omega_D} d\omega\alpha_{(t)}^2 F(\omega)[2n_B(\omega) + n_F(\omega - \varepsilon_k) + n_F(\omega + \varepsilon_k)] \tag{8.149}$$

High temperature is defined as T greater than the Debye temperature. Most solids have Debye temperatures less than room temperature, so that $T = 300$ K is a high temperature. In this limit, the largest term in the bracket is $n_B(\omega) \approx k_B T/\hbar\omega$ and the lifetimes have the simple expression

$$\frac{\hbar}{\tau_p} = 2\pi\lambda k_B T \tag{8.150}$$

$$\frac{\hbar}{\tau_{pt}} = 2\pi\lambda_t k_B T \tag{8.151}$$

$$\lambda_{(t)} = 2\int_0^{\omega_D} \frac{d\omega}{\omega}\alpha_{(t)}^2 F(\omega) \tag{8.152}$$

The inverse lifetime is proportional to temperature, and the constant of proportionality is the dimensionless electron–phonon coupling constant $\lambda_{(t)}$. The resistivity of metals $[\rho = m/(e^2 n_0 \tau_t)]$ at high temperature is linear in T, and the slope is given by λ_t. A measurement of $\rho(T)$, gives λ_t, which can then be used to predict the transition temperature for superconductivity in that metal. This process works quite well.

The equation for energy relaxation is

$$\left(\frac{dE}{dt}\right)_{pe} = 2\pi\hbar\int_0^{\omega_D} d\omega\omega\alpha^2 F(\omega)[1 - n_F(\varepsilon_k - \omega) - n_F(\varepsilon_k + \omega)]$$

$$\approx [1 - 2n_F(\varepsilon_k)]\hbar\pi\lambda\langle\omega^2\rangle \tag{8.153}$$

$$\lambda\langle\omega^2\rangle = 2\int_0^{\omega_D} d\omega\omega\alpha^2 F(\omega) \tag{8.154}$$

The frequency integrals go from zero to the Debye frequency $\omega_D = k_B\Theta_D/\hbar$. The occupation numbers were expanded $n_F(\varepsilon_k \pm \omega) \approx n_F(\varepsilon_k) \pm \omega n_F'$. The first derivative terms cancel and the second derivative terms are small and are neglected. The factor $[1 - 2n_F(\varepsilon_k)]$ is one for electrons and minus one for holes. The energy relaxation always takes the particle to the chemical potential.

The energy relaxation (dE/dt) is governed by the quantity $\lambda\langle\omega^2\rangle$. It is a single function, although the notation gives the impression that it is a product of two functions. The quantity $\hbar\lambda\langle\omega^2\rangle$ has the units of Watts. In this case there is no obvious lifetime τ_{pe} nor mfp (l_{pe}). The phonons give a constant value to the energy relaxation as long as the electron's energy $\xi = \varepsilon_k - \mu > \hbar\omega_D$.

8.3.2. Semiconductors

The relaxation time for a semiconductor is calculated assuming that there is only one electron in the band. Most semiconductors have electrons or hole in one or several band minimum. In order to keep the discussion simple, the present calculation will be done

assuming there is a single conduction band at the center of the Brillouin zone. This situation applies to GaAs and other III–V and II–VI semiconductors. The phonon wave vectors are rather small. An electron with energy $\varepsilon_k \approx k_B T$ has a small wave vector k. If it emits or absorbs a phonon and goes to energy $\varepsilon_{\mathbf{k}+\mathbf{q}} \pm \hbar\omega_q$, then $|\mathbf{k}+\mathbf{q}|$ is a small wave vector. The consequence is that q is also a small wave vector. At small wave vectors, it is a good approximation to represent acoustical phonons by the Debye model ($\omega_q = c_s q$) and optical phonons by an Einstein model ($\omega_q = \omega_0$). The results for acoustical phonons are derived here. The case of optical phonons is assigned as problems, although the derivation is similar to that for Fröhlich polarons in the last section.

The wave vector integrals can be done analytically since the Debye approximation is accurate. Define

$$\alpha^2 F^{(\pm)}(k, \omega) = \int \frac{d^3 q}{(2\pi)^3} |M_q|^2 \delta(\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}} \pm \hbar\omega_{\mathbf{q}}) \delta(\omega - \omega_{\mathbf{q}}) \tag{8.155}$$

$$= \frac{D^2 m^*}{8\pi^2 \hbar^2 \rho k c_s^4} \omega^2 |_{\omega_l < \omega < \omega_u} \tag{8.156}$$

$$\omega_u = 2c_s(k \pm k_s), \qquad \omega_l = 2c_s(\pm k_s - k) \tag{8.157}$$

$$k_s = m^* c_s / \hbar \tag{8.158}$$

The lower limit ω_l can be set equal to zero since the factor of ω^2 makes this contribution negligible. The lifetime is

$$\frac{1}{\tau_p(k)} = 2\pi \int d\omega \{\alpha^2 F^{(-)}(k, \omega)[n_B(\omega) + 1] + \alpha^2 F^{(+)}(k, \omega) n_B(\omega)\}$$

There are two interesting limits to this expression. The first is at zero temperature where the phonon occupation numbers $n_B(\omega) = 0$. The particle is assumed to be energetic so that $k \gg k_s$. In this case the answer is

$$\frac{1}{\tau_p(k)} = \gamma_p \varepsilon_k \tag{8.159}$$

$$\gamma_p = \frac{4}{3\pi} \frac{D^2 m^{*2}}{\hbar^4 \rho c_s} \tag{8.160}$$

The lifetime depends upon the value of kinetic energy. The second case is at high temperature, where $n_B \approx k_B T / \hbar\omega$ so the result is proportional to temperature. In this case the natural quantity is the mfp $[l_p(k) = v_k \tau_p(k)]$

$$\frac{1}{\tau_p(k)} = \frac{v_k}{l_p} \tag{8.161}$$

$$\frac{1}{l_p} = \frac{D^2 k_B T m^{*2}}{\pi \hbar^4 \rho c_s^2} \tag{8.162}$$

Another mfp for the electron is called l_ϕ. The symbol ϕ denotes the phase of the electron. The value of l_ϕ is the distance over which the electron travels before it breaks its phase coherence. For electrons in a pure semiconductor, where there are no electron–electron interactions, then $l_\phi = l_p$. The distance for phase coherence is given by the mfp for scattering by phonons. The phonon can carry away an arbitrary amount of phase, so such scattering does change the

phase of the electron in a random fashion. In contrast, the scattering by impurities does not change the phase coherence. Impurity scattering changes the phase of the electron, but it changes the phase of each electron by the same amount. So coherence is maintained in scattering by impurities, while it is not maintained in the scattering by phonons. The impurities are just part of the one-electron potential which guides the electron as it wanders through the crystal. Electron–electron interactions also break phase coherence.

A calculation similar to finding l_p can be done for the momentum relaxation

$$\alpha_t^2 F^{(\pm)}(\omega) = \int \frac{d^3q}{(2\pi)^3} |M_q|^2 (-\mathbf{k}\cdot\mathbf{q}/k^2)\delta(\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}} \pm \hbar\omega_{\mathbf{q}})\delta(\omega - \omega_{\mathbf{q}})$$

$$= \frac{D^2 m^*}{16\pi^2\hbar^2\rho k^3 c_s^6}\omega^4|_{\omega_l<\omega<\omega_u}$$

$$\frac{1}{\tau_{pt}(k)} = 2\pi\int_0^{\omega_D} d\omega\{\alpha_t^2 F^{(-)}(\omega)[n_B(\omega)+1] + \alpha_t^2 F^{(+)}(\omega)n_B(\omega)\}$$

The results are at low temperature

$$\frac{1}{\tau_{pt}(k)} = \gamma_{pt}\varepsilon_k \tag{8.163}$$

$$\gamma_{pt} = \frac{8}{5\pi}\frac{D^2 m^{*2}}{\hbar^4\rho c_s} = \tfrac{6}{5}\gamma_p \tag{8.164}$$

and at high temperature

$$\frac{1}{l_{pt}} = \frac{D^2 k_B T m^{*2}}{\pi\hbar^4\rho c_s^2} = \frac{1}{l_p} \tag{8.165}$$

The results are the same at high temperatures for the transport and regular scattering mfp: $l_{pt} = l_p$. This equality is expected for the following reason. At high temperatures, the wave vector dependence of the interaction is effectively

$$\frac{|M_{\mathbf{q}}|^2}{\hbar\omega_{\mathbf{q}}} = \frac{D^2 q^2}{2\rho\omega_q^2} = \frac{D^2}{2\rho c_s^2} \tag{8.166}$$

The interaction is independent of wave vector, so the scattering is isotropic. The angular factor of $\cos(\theta)$ averages to zero for isotropic scattering, and the two mfp's are identical.

Some numerical results are presented in Table 8.1 for the high-temperature case using $T = 300\,\mathrm{K}$. The factor of $\rho c_s^2 = C_{11}$, where C_{11} is the elastic constant for LA phonons. The values for l_{pt} are several microns.

Also included are data for the polar scattering by optical phonons, which produces a lifetime τ_0, which can be converted to a mfp using the thermal velocity $v_T = \sqrt{2k_B T/m^*}$.

Energy relaxation in semiconductors is discussed using the formula (8.144)

$$\left(\frac{dE}{dt}\right) = -2\pi\int_0^{\omega_D} d\omega\omega\{\alpha^2 F^{(-)}(k,\omega)[n_B(\omega)+1] - \alpha^2 F^{(+)}(k,\omega)n_B(\omega)\}$$

At low temperature the phonon occupation number $n_B = 0$ and the only process is phonon emission. Assuming that the particle has enough energy to emit phonons, the above formula gives

$$\left(\frac{dE}{dt}\right) = \frac{D^2 m^* k^3}{\pi \hbar^2 \rho} = -\frac{v_k \varepsilon_k}{l_{pe}} \tag{8.167}$$

$$\left(\frac{dE}{dx}\right) = \frac{1}{v_k}\left(\frac{dE}{dt}\right) = -\frac{\varepsilon_k}{l_{pe}} \tag{8.168}$$

$$\frac{1}{l_{pe}} = \frac{2D^2 m^{*3}}{\pi \hbar^4 \rho} \tag{8.169}$$

The mfp path for energy relaxation l_{pe} from phonons is quite long. Some values are shown in Table 8.1. Typical distances are millimeters. Most semiconductor devices are much smaller than these length scales, so that electrons do not achieve energy relaxation in traversing most devices unless there are sufficient impurities to cause them to random walk this distance.

The other interesting case is for high temperature. Expand $n_B = k_B T/\omega_q - \frac{1}{2}$ and find

$$\left(\frac{dE}{dt}\right) = -2\pi \int_0^{\omega_D} d\omega\omega \left\{ \frac{k_B T}{\hbar\omega}[\alpha^2 F^{(-)} - \alpha^2 F^{(+)}] + \tfrac{1}{2}[\alpha^2 F^{(+)} + \alpha^2 F^{(+)}] \right\}$$

The last term gives the same mfp l_{pe} as was found for zero temperature. The first term, which is proportional to $k_B T$, is nonzero when the limits to the frequency integrals are taken to be $2c_s(k \pm k_s)$, so the integral is

$$\int^{2c_s(k-k_s)} d\omega\omega^2 - \int^{2c_s(k+k_s)} d\omega\omega^2 = \tag{8.170}$$

$$-\tfrac{1}{3}(2c_s)^3[(k+k_s)^3 - (k-k_s)^3] = -16k^2 m^* c_s^4 \tag{8.171}$$

$$\left(\frac{dE}{dx}\right) = -\frac{1}{l_{pe}}[\varepsilon_k - 2k_B T] \tag{8.172}$$

The energy relaxation has the same formula l_{pe} for the mfp at room temperature as at low temperature. The only change at higher temperature is that the energy ε_k relaxes to the value $2k_B T$, which was zero at zero temperature.

TABLE 8.1 Electron mfp data at $T = 300$ K. The top lines have material constants. The second group of lines has estimated mpf from optical phonon scattering. The last two lines are from deformation potential scattering by LA phonons

	Units	GaAs	InP	InAs	InSb
m^*	m_e	0.064	0.078	0.027	0.013
ε_0		12.8	12.5	15.15	17.7
ε_∞		10.9	9.5	12.25	15.7
C_{11}	GPa	118	101	83	66.7
ρ	g/cm^3	5.32	4.81	5.67	5.78
$\hbar\omega_{LO}$	meV	35.4	42.8	29.6	23.6
D	eV	8.0	6.4	6.0	14.0
α		0.067	0.125	0.055	0.020
τ_0	ps	0.80	0.51	0.85	2.09
v_T	km/s	378	341	581	837
$v_T\tau_0$	μm	0.30	0.18	0.50	1.75
l_{pt}	μm	2.0	1.8	13.9	8.8
l_{pe}	mm	3.2	2.5	80	134

8.3.3. Temperature Relaxation

The relaxation time for temperature is required when the electron temperature T_e and the phonon temperature T_p are different. The electron–phonon interaction will allow energy exchange between these two systems, and gradually bring them to the same temperature. The rate at which they equilibrate defines the temperature relaxation time. The result for metals was introduced by Allen (1987). The same quantity is important in semiconductors. There are several situations where an experimentalist might encounter $T_p \neq T_e$. Since the electrons absorb electromagnetic radiation, an intense laser pulse could raise T_e above T_p. Similarly, a strong dc electric field could accelerate electrons to have an average kinetic energy well above the phonon temperature, and then the two systems would mutually relax to the same temperature.

The starting expression for this calculation is the collision term (8.137) in the Boltzmann equation. The expression is simplified with the following assumptions:

- The electrons are in thermal equilibrium among themselves with an effective temperature T_e, $\beta_e = 1/k_B T_e$ and distribution function $f(\mathbf{p}) = n_F(\xi_{\mathbf{p}})$. This relative equilibrium is maintained by rapid electron–electron scattering.
- The phonons are in thermal equilibrium among themselves with an effective temperature T_p, $\beta_p = 1/k_B T_p$ which is maintained by rapid anharmonic phonon–phonon scattering. They are described by a distribution function $n_B(\omega_{\mathbf{q}})$.
- The electrons and phonons exchange energy according to (8.137).

Always keep in mind that the boson distribution function $n_B(\omega)$ is at a different temperature than the electron distribution functions $n_F(\varepsilon)$, $n_F(\varepsilon')$. The expression can be further simplified by using the function $\alpha^2 F(\omega)$ to express the integrals over the phonon states

$$\begin{aligned}\left(\frac{\partial f(\varepsilon)}{\partial t}\right)_{ep} &= 2\pi \int d\omega \alpha^2 F(\omega) \int d\varepsilon' \\ &\times \{\delta(\varepsilon - \varepsilon' - \omega)[n_F(\varepsilon)[1 - n_F(\varepsilon')][n_B(\omega) + 1] - n_F(\varepsilon')n_B(\omega)[1 - n_F(\varepsilon)]] \\ &+ \delta(\varepsilon - \varepsilon' + \omega)[n_F(\varepsilon)n_B(\omega)[1 - n_F(\varepsilon')] - n_F(\varepsilon')(n_B(\omega) + 1)[1 - n_F(\varepsilon)]]\}\end{aligned} \tag{8.173}$$

where $\varepsilon = \varepsilon_{\mathbf{p}}$, $\varepsilon' = \varepsilon_{\mathbf{p}+\mathbf{q}}$. Next, the change in the internal energy is calculated. The change in energy is accomplished by inserting under in the integral the factor of $\varepsilon' - \varepsilon = \pm\hbar\omega$.

$$\begin{aligned}\left(\frac{\partial E(\varepsilon)}{\partial t}\right)_{ep} &= -2\pi\hbar \int d\omega \alpha^2 F(\omega)\omega \int d\varepsilon' \quad (8.174) \\ &\times \{\delta(\varepsilon - \varepsilon' - \omega)[n_F(\varepsilon)[1 - n_F(\varepsilon')][n_B(\omega) + 1] - n_F(\varepsilon')n_B(\omega)[1 - n_F(\varepsilon)]] \\ &- \delta(\varepsilon - \varepsilon' + \omega)[n_F(\varepsilon)n_B(\omega)[1 - n_F(\varepsilon')] - n_F(\varepsilon')(n_B(\omega) + 1)[1 - n_F(\varepsilon)]]\} \\ &= -2\pi\hbar \int d\omega \alpha^2 F(\omega)\omega \\ &\times \{n_F(\varepsilon)[1 - n_F(\varepsilon - \omega)][n_B(\omega) + 1] - n_F(\varepsilon - \omega)n_B(\omega)[1 - n_F(\varepsilon)] \\ &- n_F(\varepsilon)n_B(\omega)[1 - n_F(\varepsilon + \omega)] + n_F(\varepsilon + \omega)(n_B(\omega) + 1)[1 - n_F(\varepsilon)]\}\end{aligned} \tag{8.175}$$

In order to average over the entire system of electrons, it is necessary to also integrate over $d\varepsilon$. If $N(\varepsilon)$ is the density of states for the electron system, the energy change in the electron system is

$$\frac{dU}{dt} = C\frac{dTe}{dt} = \int d\varepsilon N(\varepsilon)\left(\frac{\partial E(\varepsilon)}{\partial t}\right)_{ep} \tag{8.176}$$

where C is the heat capacity. The integrals over $d\varepsilon$ all converge within a thermal energy of the chemical potential. On this small energy scale, the density of states $N(\varepsilon)$ can be taken to be a constant $N(0)$, where the zero of energy is the chemical potential. Then all of the integrals over $d\varepsilon$ have the typical form

$$I(\omega) = \int d\varepsilon n_F(\varepsilon)[1 - n_F(\varepsilon - \omega)] = \frac{\omega}{e^{\beta_e\omega} - 1} = \omega n_{Be}(\omega) \tag{8.177}$$

The subscript "e" is added to the boson occupation function $n_{Be}(\omega)$ to emphasize that it is evaluated at the electron temperature T_e. The above result can be derived by changing variables of integration to $s = \exp(\beta_e\varepsilon)$, $ds = s\beta_e d\varepsilon$, $t = \exp(\beta_e\omega)$ which changes the integral to

$$I(\omega) = k_B T_e \int_0^\infty \frac{ds}{(s+1)(s+t)} = \frac{k_B T_e}{e^{\beta_e\omega} - 1}\ln\left[\frac{s+1}{s+t}\right]_0^\infty = \omega n_{Be}(\omega)$$

The other integrals over $d\varepsilon$ are evaluated in a similar fashion. The result for the energy relaxation is

$$C\frac{dT_e}{dt} = -4\pi\hbar^2 N(0)\int_0^{\omega_D} d\omega\omega^2\alpha^2 F(\omega)[n_{Be}(\omega) - n_{Bp}(\omega)] \tag{8.178}$$

The final formula is quite simple. The right-hand side of this equation obviously vanishes in equilbrium when $T_e = T_p$. Another simplification occurs because the heat capacity is proportional to the density of states $C = \pi^2 k_B^2 TN(0)/3$ which simplifies the expression to

$$\frac{dT_e}{dt} = -\frac{12\hbar^2}{\pi k_B^2 T_e}\int_0^{\omega_D} d\omega\omega^2\alpha^2 F(\omega)[n_{Be}(\omega) - n_{Bp}(\omega)] \tag{8.179}$$

The only factor which relates to the particular solid is $\alpha^2 F(\omega)$.

There are two interesting limits when evaluating this expression. The first is at very low temperature, $T_{e,p} \ll \Theta$, where the Debye temperature is Θ. Then the integral over ω has its main contribution at small values of ω. The small values of frequency come from sound waves, where it is suitable to use the Debye approximation. The evaluation of $\alpha^2 F(\omega)$ in this limit is identical to the semiconductor case, which gives that $\alpha^2 F = \gamma\omega^2$ where γ is a constant. The frequency limits can be extended to infinity, which gives the expression

$$\frac{dT_e}{dt} = -\Gamma[T_e^5 - T_p^5] \tag{8.180}$$

$$\Gamma = \frac{12k_B^3\gamma}{\pi T_e\hbar^3} I_4 \tag{8.181}$$

$$I_4 = \int_0^\infty \frac{dxx^4}{e^x - 1} = 4!\zeta(5) \tag{8.182}$$

The relaxation obeys a T^5 law, which means very little heat is exchanged between electrons and phonons at low temperatures.

The other interesting limit is at room temperature, or at least above the Debye temperature. Then the boson occupation factors can be expanded $n_B \approx k_B T/\hbar\omega$ which gives the simple result

$$\frac{dT_e}{dt} = -\frac{(T_e - T_p)}{\tau_{pT}} \tag{8.183}$$

$$\frac{1}{\tau_{pT}} = \frac{6\hbar}{\pi k_B T_e}\lambda\langle\omega^2\rangle \tag{8.184}$$

At high temperature the temperature relaxation is determined by the temperature difference $\delta T = T_e - T_p$, as well as by the function $\lambda\langle\omega^2\rangle$. The latter function also determines the energy relaxation.

The temperature relaxation is measured by lasers using pulse-probe techniques. The first laser pulse is absorbed by the electrons, which thermally excites them, which causes their temperature to differ from the phonons. The time-delayed probe measurement determines how some property, such as the refractive index, varies with the time interval after the initial pulse. An observation of an exponential relaxation curve is a measurement of τ_{pT}.

The above result is valid for a metal. A similar derivation can be done for a semiconductor, which is assigned in the problems. The same calculation for scattering of electrons in a semiconductor, by optical phonons, can be done in a interesting and rigorous way. For optical phonon scattering in a semiconductor, all of the phonons have the same energy $\hbar\omega_0$ so there is no need for an integral over frequency. In fact, just set

$$\alpha^2 F(\omega) = \frac{\lambda\omega_0}{2}\delta(\omega - \omega_0) \tag{8.185}$$

where λ is the dimensionless coupling constant. Starting from (8.173) gives

$$\begin{aligned}\left(\frac{\partial f(\varepsilon)}{\partial t}\right) = -\pi\lambda\omega_0\{[f(\varepsilon)(N_0+1) - N_0 f(\varepsilon-\omega_0)]\Theta(\varepsilon-\omega_0)\\ + f(\varepsilon)N_0 - f(\varepsilon+\omega_0)(N_0+1)\}\end{aligned} \tag{8.186}$$

The states $f(\varepsilon + l\omega_0)$ can be regarded as a ladder of energy states, where each level is separated by ω_0. The above equation can be solved exactly by matrix methods. Define $f_l = f(\varepsilon + l\omega_0)$ where the energy ε is now limited to the range $0 < \varepsilon < \omega_0$. The vector $\tilde{f}$ has f_l as its elements. The above equation can now be cast into a matrix equation, where N_0 is the occupation number of the optical phonons.

$$\frac{d}{dt}\tilde{f} = -\frac{1}{t_{oT}}\mathscr{M}\cdot\tilde{f} \tag{8.187}$$

$$\mathscr{M} = \begin{pmatrix} N_0 & -(N_0+1) & 0 & \cdots \\ -N_0 & 2N_0+1 & -(N_0+1) & \cdots \\ 0 & -N_0 & 2N_0+1 & \cdots \end{pmatrix} \tag{8.188}$$

$$\frac{1}{\tau_{oT}} = \pi\lambda\omega_0 \tag{8.189}$$

The matrix $\mathscr{M}$ has tridiagonal form. Only three rows are shown, but its dimensionality is infinite. All row except the first have elements: $-N_0$, $2N_0 + 1$, $-(N_0 + 1)$ to the left of the

diagonal, on the diagonal, and to the right of the diagonal. If the matrix $\mathscr{M}$ has eigenvalues ε_j and eigenfunctions ψ_j then the solution is

$$\tilde{f}(t) = \sum_j a_j \psi_j e^{-\varepsilon_j t/\tau_{0T}} \tag{8.190}$$

where the coefficients a_j are determined by the initial conditions. The equilibrium distribution is given by $\varepsilon_0 = 0$, $\psi_0 = \exp(-l\beta\omega_0)$, where β is determined by the phonon temperature. The system relaxes towards the equilibrium distribution, which is why its eigenvalue has to be zero. The other eigenvalues and eigenvectors are given by Mahan (1985).

8.4. ELECTRON–PHONON INTERACTIONS IN METALS

8.4.1. Force–Force Correlation Function

In pure metals the electrical resistivity has two components. There is usually a constant resistivity from electron scattering by impurities, which is the largest part of the resistance at small temperatures. There is also a resistivity from electron scattering by phonons, which is temperature dependent and becomes large at high temperature. "Matthiessen's rule" (1862) is that these two contributions to the resistance are additive. It should be regarded as a rule of thumb, rather than an ironclad rule. There are enough "deviations from Matthiessen's rule" to make the abbreviation DMR a familiar acronym (Bass, 1982).

Electron scattering by acoustical phonons presents a hard problem in transport theory. The scattering is slightly inelastic. This problem is solved using neither the elastic scattering theory of Sec. 8.1 nor the inelastic scattering theory of Sec. 8.2. Instead an integral equation must be derived and solved for the energy dependence of the scattering process. The slightly inelastic nature of the scattering process makes this calculation much harder than the previous cases.

Two methods for obtaining the electrical conductivity are emphasized in this book. One uses equilibrium methods and evaluates the Kubo formula for the current–current correlation function. The resistivity from phonons will be found using this method, which follows the original derivation by Holstein (1964). The second method utilizes the quantum Boltzmann equation (QBE), which is a nonequilibrium theory. Mahan and Hansch (1983) used the QBE to derive the Holstein formula. Both of these derivations are complicated. They end by deriving the same integral equation for the scattering function, which must be solved by further work. Their virtue is that they are formally exact starting points, although approximations are made in obtaining the solution.

Other methods for obtaining the resistivity have been proposed, partly to avoid all of the work associated with the exact methods. These other methods are approximate. However, often the theories are both simple and accurate, which make them useful approximations. One of them is the force–force correlation function. If $\mathbf{F}(t)$ is the fluctuating force that acts on the electron, then define $R(i\omega)$ as the force–force correlation function:

$$R(i\omega) = -\frac{1}{3}\int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau \mathbf{F}(\tau) \cdot \mathbf{F}(0) \rangle \tag{8.191}$$

$$\rho = \frac{1}{e^2 n_0^2} \lim_{\omega \to 0} \left[\frac{\operatorname{Im} R_{\text{ret}}(\omega)}{\omega} \right] \tag{8.192}$$

This formula is just the quantum analogy of the Nyqvist theorem (Kubo *et al.*, 1985). After calculating this correlation function, the retarded function is obtained by letting $i\omega \to \omega + i\delta$. The resistance ρ is found by dividing by ω and taking the limit of $\omega \to 0$. For example, assume that the force on the electron has two terms: $\mathbf{F}_i$ from impurities and $\mathbf{F}_{ph}$ from phonons. If they are uncorrelated, then the correlation function has no cross terms. Symbolically write

$$R = \langle(\mathbf{F}_i + \mathbf{F}_{ph}) \cdot (\mathbf{F}_i + \mathbf{F}_{ph})\rangle = \langle \mathbf{F}_i \cdot \mathbf{F}_i\rangle + \langle \mathbf{F}_{ph} \cdot \mathbf{F}_{ph}\rangle \tag{8.193}$$

In this case the resistivities from impurities and phonons are additive, in agreement with Matthiessen's rule.

As an example, the resistivity is calculated from impurity scattering. The potential energy of the electron scattering from the impurities at $\mathbf{R}_i$ is discussed in Sec. 4.1.5

$$V(\mathbf{r}) = \sum_i V_{ei}(\mathbf{r} - \mathbf{R}_i) = \frac{1}{\mathrm{v}} \sum_{i\mathbf{q}} V(\mathbf{q}) \exp[i\mathbf{q} \cdot (\mathbf{r} - \mathbf{R}_i)] \tag{8.194}$$

The force $\mathbf{F}$ is the gradient of the potential. The factor of $\exp(i\mathbf{q} \cdot \mathbf{r})$ can also be written as the electron density operator $\rho(\mathbf{q})$. The factor of $\exp(-i\mathbf{q} \cdot \mathbf{R}_i)$ can be written as the impurity density operator $\rho_i(-\mathbf{q})$

$$\mathbf{F}(\mathbf{r}) = -\frac{i}{\mathrm{v}} \sum_{i\mathbf{q}} \mathbf{q} V(\mathbf{q}) \exp[i\mathbf{q} \cdot (\mathbf{r} - \mathbf{R}_i)] = -\frac{i}{\mathrm{v}} \sum_{i\mathbf{q}} \mathbf{q} V(\mathbf{q}) \rho(\mathbf{q}) \rho_i(-\mathbf{q})$$

The next step is to evaluate the force–force correlation function $R(i\omega)$. In correlating $\mathbf{F}$ with itself, there are two separate factors. One is $\langle \rho_i(-\mathbf{q}) \rho_i(-\mathbf{q}')\rangle$, which equals the number of impurities N_i if $\mathbf{q} = -\mathbf{q}'$. The other is the electron density–density correlation function, which is given exactly in terms of the inverse dielectric function

$$R_i(i\omega) = -\frac{N_i}{3\mathrm{v}^2} \sum_{\mathbf{q}} \frac{q^2}{v_q} V(\mathbf{q})^2 \left[\frac{1}{\varepsilon(\mathbf{q}, i\omega)} - 1\right] \tag{8.195}$$

The next step is to take the imaginary part of the retarded function. The only retarded function on the right-hand side of the above equation is the inverse dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$. Its imaginary part is $-\varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2)$. At low frequency $\varepsilon_2 = 2\omega e^2 m^2/q^3$. Dividing by ω eliminates the factor of frequency. The formula for the resistivity from impurity scattering is

$$\rho = \frac{n_i m^2}{6\pi n_0^2 e^2} \int \frac{d^3 q}{(2\pi)^3} q \left| \frac{V(q)}{\varepsilon(q)} \right|^2 \tag{8.196}$$

This formula is the exact result for the zero-temperature resistivity from impurity scattering, when the scattering is calculated in the second Born approximation. If $V(q)/\varepsilon(q)$ is replaced by the T matrix for scattering, then it is the exact result, period. It is the formula $\rho = m/(n_0 e^2 \tau_t)$, where τ_t is defined in (8.25). It even includes the factor of $(1 - \cos\theta')$, although this assertion is not immediately obvious. One has to perform the angular integral in (8.25), which eliminates the delta function, in order to show its equivalence with the above formula for the resistivity.

The force–force correlation function gives the right resistivity for impurity scattering. No vertex correction or integral equation was needed in the derivation. The ease of derivation has made this approach popular.

Several caveats are needed. One is that impurity scattering is the only known example where the force–force correlation function gives the correct answer. In other cases it give an approximate answer. The second caveat is that the right answer is obtained by a wrong

derivation. The derivation contains two important limits. One is setting the volume $v \to \infty$, while the second is $\omega \to 0$. The above answer is obtained by taking these limits in the wrong order. If they had been done correctly, in the right order, a different answer is obtained. Using the right order in evaluating the force–force correlation function gives a zero result as $\omega \to 0$. These points are discussed by Argyres and Sigel (1974), Huberman and Chester (1975), Kubo *et al.* (1985), and Fishman (1989).

The force–force correlation function may also be evaluated for the electron scattering by phonons. The result is

$$\rho(T) = C' \sum_{\lambda} \int q d^3q |W(\mathbf{q})|^2 (\hat{\xi}_\lambda \cdot \mathbf{q})^2 \left[-\frac{dn_B(\omega)}{d\omega} \right] \tag{8.197}$$

$$C' = \frac{3\hbar v_0}{16 M e^2 v_F^2 k_F^4} \tag{8.198}$$

Here $W(\mathbf{q})$ is the screened electron–ion interaction and v_0 is the unit cell volume. This formula was first derived by Ziman (1960) as a variational solution to the Boltzmann equation. It is the formula that is most often evaluated when calculating the temperature dependence of the resistivity of metals. Figure 8.8 shows a theoretical calculation of Dynes and Carbotte (1968) compared with experiments for Na and K. An important feature of these calculations is numerically integrating over the Brillouin zone for all the phonon states, while employing accurate values for the phonon frequencies $\omega_\lambda(\mathbf{q})$ and polarization vectors $\hat{\xi}$.

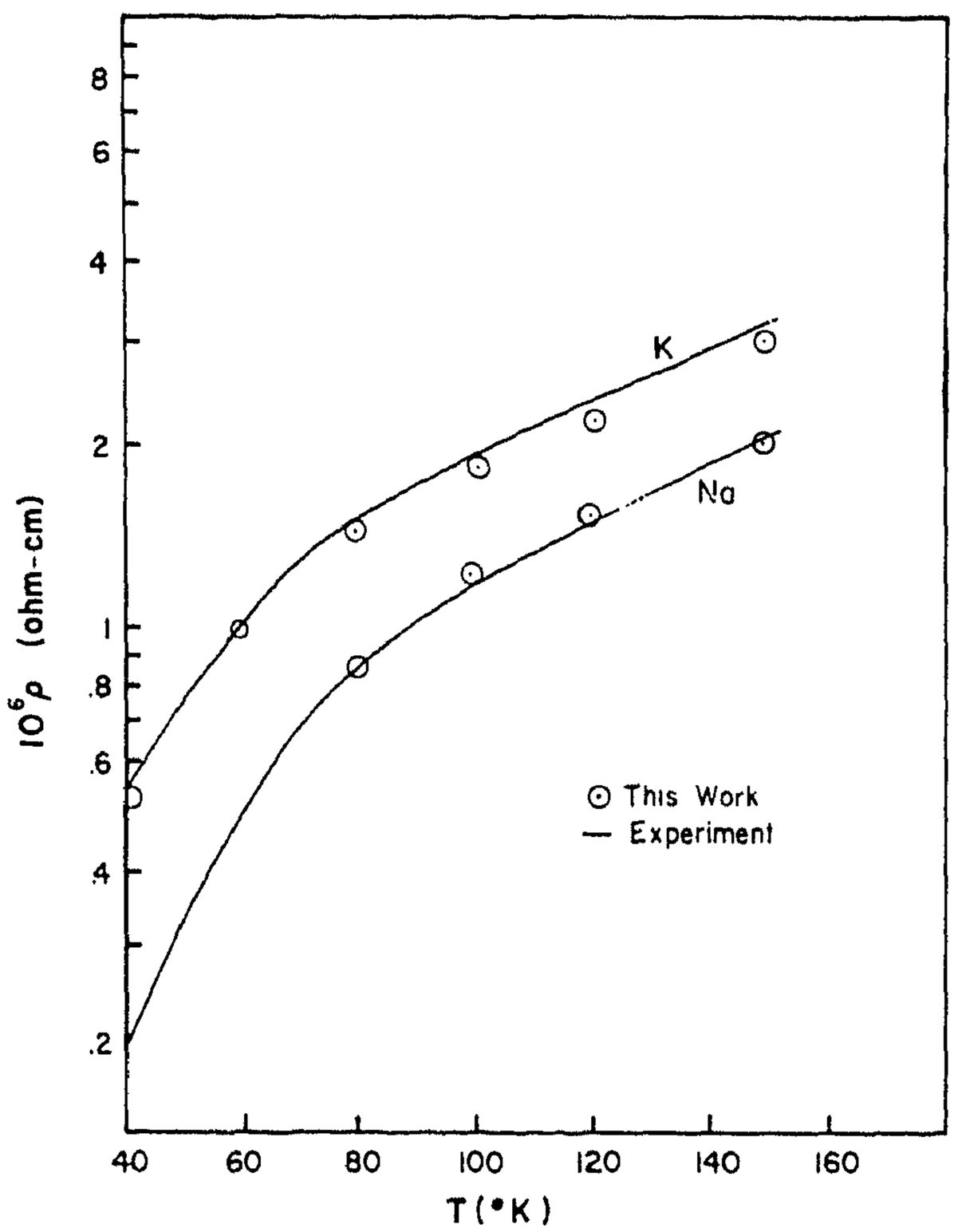


FIGURE 8.8 Resistivity as a function of temperature for Na and K. The solid line is experiment, and points are theory. *Source*: Dynes and Carbotte (1968) (used with permission).

8.4.2. Kubo Formula

A rigorous theory of the dc electrical conductivity is calculated for the scattering of electrons by all phonons: acoustical and optical. The starting point is the Kubo formula, and the derivation follows the original one by Holstein (1964). His theory sums the ladder diagrams for phonons and reduces the vertex function to an integral equation which is solved numerically. So far a solution is available only for a spherical Fermi surface, so that the result is the Kubo formula analogy of (8.198). The results are expressed in terms of the McMillan function $\alpha^2F(\omega)$ and a similar function $\alpha_t^2F(\omega)$, which is used in transport theory.

The goal is to evaluate the current–current correlation function in the presence of the electron–phonon interaction. This correlation function can always be expressed as a product of two Green's functions and the vertex function. The Green's function $\mathcal{G}(\mathbf{p}, ip)$ used here represents fully interacting particles, with a self-energy found from the electron–phonon interaction plus any additional interactions of interest. The Feynman diagram for the correlation function is shown in Fig. 8.9 where the vertex function is put only at one end of the bubble in order not to overcount the vertex terms:

$$\begin{aligned}\pi(i\omega) &= -\frac{1}{3\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \mathbf{j}(\tau)\cdot\mathbf{j}(0)\rangle \\ &= -\frac{e^2}{3m^2\nu}\sum_{\mathbf{p}\mathbf{p}'\sigma\sigma'}\mathbf{p}\cdot\mathbf{p}'\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau C^\dagger_{\mathbf{p}\sigma}(\tau)C_{\mathbf{p}\sigma}(\tau)C^\dagger_{\mathbf{p}'\sigma'}(0)C_{\mathbf{p}'\sigma'}(0)\rangle \qquad (8.199)\end{aligned}$$

$$\pi(i\omega) = \frac{2e^2}{3m^2}\int\frac{d^3p}{(2\pi)^3}\frac{1}{\beta}\sum_{ip}\mathcal{G}(\mathbf{p}, ip)\mathcal{G}(\mathbf{p}, ip+i\omega)\mathbf{p}\cdot\mathbf{\Gamma}(\mathbf{p}; ip, ip+i\omega) \qquad (8.200)$$

The dc conductivity is found by the same steps used in Secs. 8.1 and 8.2. One evaluates the correlation function for values of $i\omega$ and analytically continues $i\omega \to \omega + i\delta$ to find the retarded function. The dc conductivity is the imaginary part of the retarded function divided by ω, in the limit where $\omega \to 0$.

The vertex function $\mathbf{\Gamma}(\mathbf{p}; ip, ip+i\omega)$ is evaluated below. It depends on both the frequency variables ip and $i\omega$, which is written in the combination $(ip, ip+i\omega)$. The two arguments ip and $ip+i\omega$ come from the electron Green's functions which have the same frequency arguments. In a homogeneous electron gas, the vector vertex function $\mathbf{\Gamma}(\mathbf{p}; ip, ip+i\omega)$ must point in the vector direction $\mathbf{p}$, although in real metals the crystalline potential defines other possible directions. However, the approximation of treating the Fermi surface as strictly spherical is equivalent to neglecting crystal directions, so assume $\mathbf{\Gamma}$ points

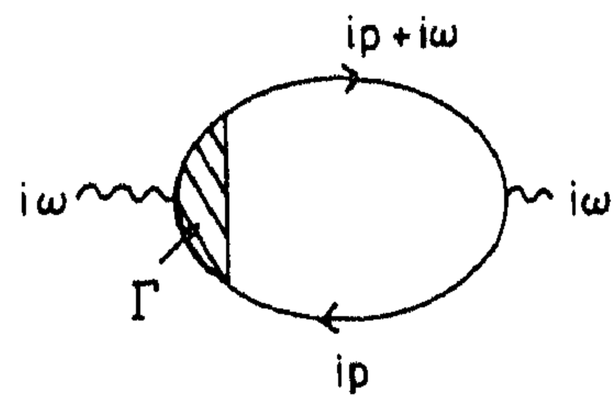


FIGURE 8.9

in the direction $\mathbf{p}$. It is convenient to introduce the scalar function $\gamma(\mathbf{p}; ip, ip + i\omega)$, which is the amplitude of the vector vertex function:

$$\Gamma(\mathbf{p}; ip, ip + i\omega) = \mathbf{p}\gamma(\mathbf{p}; ip, ip + i\omega) \tag{8.201}$$

$$\pi(i\omega) = \frac{2e^2}{3m^2}\int \frac{d^3p}{(2\pi)^3} p^2 S(\mathbf{p}, i\omega) \tag{8.202}$$

$$S(\mathbf{p}, i\omega) = \frac{1}{\beta}\sum_{ip} \mathscr{G}(\mathbf{p}, ip)\mathscr{G}(\mathbf{p}, ip + i\omega)\gamma(\mathbf{p}; ip, ip + i\omega) \tag{8.203}$$

The scalar function γ is not the same as the scalar vertex function in (8.81).

The next step is to evaluate the summation over Matsubara frequencies ip to obtain $S(\mathbf{p}, i\omega)$. To this end, construct the usual contour integral which has cuts along the axes where $ip \to$ real and also $ip + i\omega \to$ real. These series of algebraic operations are the same as used to derive (8.64) for impurity scattering. That earlier result can be used here by identifying $P = GG\gamma$

$$\begin{aligned}\mathrm{Im}[S_{\mathrm{ret}}(\mathbf{p}, \omega)] = \mathrm{Re}\Bigg\{&\int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi}[n_F(\varepsilon + \omega) - n_F(\varepsilon)] \\ &\times [G_{\mathrm{adv}}(\mathbf{p}, \varepsilon)G_{\mathrm{ret}}(\mathbf{p}, \varepsilon + \omega)\gamma(\mathbf{p}; \varepsilon - i\delta, \varepsilon + \omega + i\delta) \\ &- G_{\mathrm{ret}}(\mathbf{p}, \varepsilon)G_{\mathrm{ret}}(\mathbf{p}, \varepsilon + \omega)\gamma(\mathbf{p}; \varepsilon + i\delta, \varepsilon + \omega + i\delta)]\Bigg\}\end{aligned} \tag{8.204}$$

and the conductivity is

$$\begin{aligned}\sigma = \frac{2e^2}{3m^2}\int \frac{d^3p}{(2\pi)^3} p^2 \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi}\left[-\frac{dn_F(\varepsilon)}{d\varepsilon}\right] \\ \times \{|G_{\mathrm{ret}}(\mathbf{p}, \varepsilon)|^2\gamma(\mathbf{p}, \varepsilon - i\delta, \varepsilon + i\delta) - \mathrm{Re}[G_{\mathrm{ret}}(\mathbf{p}, \varepsilon)^2\gamma(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta)]\}\end{aligned} \tag{8.205}$$

Equation (8.205) is exact for the exact vertex function. It is expressed in terms of the two functions $\gamma(\mathbf{p}; \varepsilon - i\delta, \varepsilon + i\delta)$ and $\gamma(\mathbf{p}; \varepsilon + i\delta, \varepsilon + i\delta)$. These two functions are expected to be quite different, as they were in Sec. 8.1 for impurity scattering. The function $\gamma(\mathbf{p}; \varepsilon + i\delta, \varepsilon + i\delta)$ could be obtained from a Ward identity. That is true here, as first shown for the electron–phonon system in metals by Englesberg and Schrieffer (1963). This Ward identity can be expressed in terms of $\Gamma(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta)$ or the scalar $\gamma(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta)$:

$$\Gamma(\mathbf{p}; \varepsilon + i\delta, \varepsilon + \omega + i\delta) = \mathbf{p} + m\nabla_{\mathbf{p}}\Sigma(\mathbf{p}, \varepsilon) \tag{8.206}$$

$$\gamma(\mathbf{p}; \varepsilon + i\delta, \varepsilon + \omega + i\delta) = 1 + \frac{\partial}{\partial\xi_p}\Sigma(\mathbf{p}, \varepsilon) \tag{8.207}$$

$$G_{\mathrm{ret}}(\mathbf{p}, \varepsilon)^2\gamma(\mathbf{p}; \varepsilon + i\delta, \varepsilon + \omega + i\delta) = \frac{\partial}{\partial\xi_p}G_{\mathrm{ret}}(\mathbf{p}, \varepsilon) \tag{8.208}$$

For metals, the electron–phonon system has the feature that the self-energy function $\Sigma(\mathbf{p}, \varepsilon)$ is not very $\mathbf{p}$ dependent, and the derivative of $\Sigma(\mathbf{p}, \varepsilon)$ with respect to ξ_p is small: $d\Sigma/d\xi_p \approx \Sigma/E_F \ll 1$. A good approximation is to set $\gamma(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta) = 1$.

There is another result which is even stronger, since it applies for all values of ε near the Fermi energy. Since the self-energy $\Sigma(\mathbf{p}, \varepsilon)$ does not have significant $\mathbf{p}$ dependence near $p \sim k_F$, call it $\Sigma(k_F, \varepsilon) = \mathrm{Re}[\Sigma(\varepsilon)] - i\Gamma(\varepsilon)$ where the p notation is suppressed. The function $\Gamma(\varepsilon) = -\mathrm{Im}[\Sigma(k_F, \varepsilon)]$. The retarded and advanced Green's function have their important dependence on p through the kinetic energy term, $\xi \equiv \xi_{\mathbf{p}}$. Define $\Omega(\varepsilon) = \varepsilon - \mathrm{Re}[\Sigma(\varepsilon)]$:

$$G_{\mathrm{ret}}(p, \varepsilon) = \frac{1}{\varepsilon - \xi - \mathrm{Re}[\Sigma(\varepsilon)] + i\Gamma(\varepsilon)} = \frac{1}{\Omega(\varepsilon) - \xi + i\Gamma(\varepsilon)} \tag{8.209}$$

$$G_{\mathrm{adv}}(p, \varepsilon) = \frac{1}{\varepsilon - \xi - \mathrm{Re}[\Sigma(\varepsilon)] - i\Gamma(\varepsilon)} = \frac{1}{\Omega(\varepsilon) - \xi - i\Gamma(\varepsilon)} \tag{8.210}$$

Similarly, the vertex function $\gamma(\mathbf{p}; \varepsilon - i\delta, \varepsilon + \omega + i\delta)$ is assumed to not have a significant dependence upon p except on the order of k_F. This assumption will be justified later. The second term in brackets in (8.205) can be neglected; it vanishes by doing the kinetic energy integration. By neglecting terms of order $O(\xi/E_F)$ then

$$d^3p p^2 = 4\pi p^4 dp = 4\pi m k_F^3 d\xi[1 + O(\xi/E_F)] = 12\pi^3 m n_0 d\xi \tag{8.211}$$

where the electron density is $n_0 = k_F^3/3\pi^2$. When the kinetic energy integral is evaluated, the only ξ variation is in the Green's functions. Of the three combinations which occur, only one makes a nonzero contribution:

$$\int_{-\infty}^{\infty} d\xi G_{\mathrm{ret}}(p, \varepsilon) G_{\mathrm{ret}}(p, \varepsilon') = 0 \tag{8.212}$$

$$\int_{-\infty}^{\infty} d\xi G_{\mathrm{ret}}(p, \varepsilon) G_{\mathrm{adv}}(p, \varepsilon') = \frac{2\pi i}{\Omega(\varepsilon) - \Omega(\varepsilon') + i[\Gamma(\varepsilon) + \Gamma(\varepsilon')]} \tag{8.213}$$

$$\int_{-\infty}^{\infty} d\xi G_{\mathrm{adv}}(p, \varepsilon) G_{\mathrm{adv}}(p, \varepsilon') = 0 \tag{8.214}$$

Each integral is evaluated by closing the contour at infinity. The two integrals which vanish have both their poles in the same half plane (upper or lower), so the integration contour can be chosen to avoid them both, which encircles no poles and hence gives zero. The integral over the combination $G_{\mathrm{adv}} G_{\mathrm{ret}}$ has one pole in each plane. Closing the contour always picks up one pole whose residue produces the result. By using these integration results, the dc conductivity (8.205) becomes

$$\sigma = \frac{e^2 n_0}{2m} \int_{-\infty}^{\infty} d\varepsilon \frac{\Lambda(\varepsilon)}{\Gamma(\varepsilon)} \left[-\frac{d}{d\varepsilon} n_F(\varepsilon) \right] \tag{8.215}$$

$$\Lambda(\varepsilon) = \gamma(k_F, \varepsilon - i\delta, \varepsilon + i\delta) \tag{8.216}$$

$$\Gamma(\varepsilon) = -2\,\mathrm{Im}[\Sigma(k_F, \varepsilon)] \tag{8.217}$$

Equation (8.215) is the final result of the formal derivation. There only remains the evaluation of the vertex function $\Lambda(\varepsilon)$ and the imaginary self-energy $\Gamma(\varepsilon)$. If the vertex function were absent ($\Lambda = 1$), then the evaluation would be easy. The imaginary self-energy $\Gamma(\varepsilon)$ has been evaluated for many metals, and a result was given in Fig. 7.14 for Pb. The quantity $1/2\Gamma = \tau(\varepsilon)$, where $\tau(\varepsilon)$ is the relaxation time defined as the average time between scattering events. The result (8.215) when $\Lambda = 1$, is just the average of the relaxation time over the thermally smeared Fermi distribution. Of course, the earlier solution for impurity scattering showed that the neglect of $\Lambda(\varepsilon)$ is a serious error. The vertex function $\Lambda(\varepsilon)$ serves the important role of weighting the scattering events and favoring those at high momentum

transfer. For impurity scattering, $\Lambda = \Gamma/\Gamma_t$ or $\Lambda/\Gamma = 1/\Gamma_t$, where Γ_t is the scattering rate which contains the equivalent of $(1 - \cos\theta')$.

The vertex equation for phonon scattering is now solved to derive Γ_t. Since the phonon scattering is inelastic, the results are not identical to those for impurity scattering. In summing the ladder diagrams for the vertex function, however, the vertex contributions are found to be important and significantly different from unity. Migdal's (1958) theorem, which asserts that vertex terms are unimportant, is contradicted.

The vertex function $\mathbf{\Gamma}(\mathbf{p}, ip, ip + i\omega)$ is calculated by solving the integral equation

$$\mathbf{\Gamma}(\mathbf{p}, ip, ip + i\omega) = \mathbf{p} + \frac{1}{\beta}\sum_{iq,\lambda}\int\frac{d^3q}{(2\pi)^3}M_\lambda(\mathbf{q})^2\mathscr{D}(\mathbf{q}, iq)\mathscr{G}(\mathbf{p}+\mathbf{q}, ip+iq)$$
$$\times\,\mathscr{G}(\mathbf{p}+\mathbf{q}, ip+iq+i\omega)\mathbf{\Gamma}(\mathbf{p}+\mathbf{q}, ip+iq, ip+iq+i\omega) \qquad (8.218)$$

This vertex sums the ladder diagrams for phonons. It is illustrated in Fig. 8.10. Iteration of Eq. (8.218) produces a series in which each additional term has one more ladder diagram. The solution to the integral equation produces an expression which contains all terms with any number of phonon ladder diagrams. This solution is not an exact evaluation of the vertex function $\mathbf{\Gamma}(\mathbf{p}, ip, ip + i\omega)$, since other vertex contributions occur which are not ladders but have the phonon lines crossed. One expects these terms to be smaller, but detailed calculations are lacking, so this is only a supposition.

It is unfortunate that the vertex function is a vector. The scalar vertex function obtained by replacing $\mathbf{p}$ by 1 in Eq. (8.218) is easily obtained from a Ward identity:

$$\Gamma(\mathbf{p}, ip, ip + i\omega) = 1 + \frac{1}{\beta}\sum_{iq,\lambda}\int\frac{d^3q}{(2\pi)^3}M_\lambda(\mathbf{q})^2\mathscr{D}(\mathbf{q}, iq)\mathscr{G}(\mathbf{p}+\mathbf{q}, ip+iq)$$
$$\times\,\mathscr{G}(\mathbf{p}+\mathbf{q}, ip+iq+i\omega)\Gamma(\mathbf{p}+\mathbf{q}, ip+iq, ip+iq+i\omega)$$
$$= 1 - \frac{1}{i\omega}[\Sigma(\mathbf{p}, ip + i\omega) - \Sigma(\mathbf{p}, ip)] \qquad (8.219)$$

The Ward identity is not helpful for our problem with a vector vertex function. The integral equation must be attacked and solved.

Define the scalar function $\gamma(\mathbf{p}, ip, ip + i\omega)$ as the scalar amplitude of the vector vertex function $\mathbf{\Gamma} = \mathbf{p}\gamma$. It obeys the integral equation

$$\gamma(\mathbf{p}, ip, ip + i\omega) = 1 + \frac{1}{\beta}\sum_{iq,\lambda}\int\frac{d^3q}{(2\pi)^3}M_\lambda(\mathbf{q})^2\mathscr{D}(q)\frac{\mathbf{p}\cdot(\mathbf{p}+\mathbf{q})}{p^2}$$
$$\times\,\mathscr{G}(p+q)\mathscr{G}(p+q+i\omega)\mathbf{\Gamma}(\mathbf{p}+\mathbf{q}, ip+iq, ip+iq+i\omega)$$

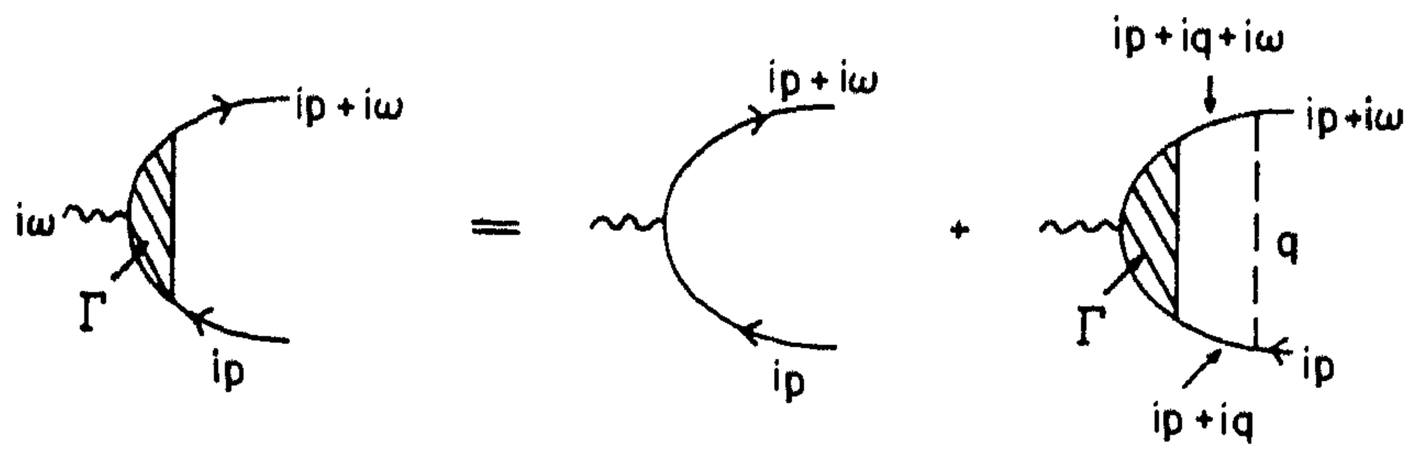


FIGURE 8.10

This equation is not the same one which is obeyed by the scalar vertex function in (8.219), and these two functions are quite different.

First do the integrals over angles and wave vector. First write $d^3q = 2\pi q^2 dq d\nu$, where $\nu = \cos\theta$ is the angle between $\mathbf{p}$ and $\mathbf{q}$. The angle variable is changed to p_1, defined as

$$p_1^2 = (\mathbf{p}+\mathbf{q})^2 = p^2 + q^2 + 2pq\nu \tag{8.220}$$

$$d\nu = \frac{p_1 dp_1}{pq} \tag{8.211}$$

$$\int d^3q = \frac{2\pi}{p}\int_0^\infty q dq \int_{|p-q|}^{|p+q|} p_1 dp_1 = \frac{2\pi}{v_0}\int_0^\infty q dq \int_{\xi_{p-q}}^{\xi_{p+q}} d\xi_1 \tag{8.222}$$

Next uncouple the limits of integration. Only electrons at the Fermi surface contribute, so that values of $\xi_1 = \varepsilon_{p_1} - \mu \approx 0$ and $\xi_p \approx 0$ are important in the integration process. Approximate $v_p = v_F$ as the Fermi velocity. Similarly, the integral qdq is understood to be over the spherical Fermi surface from one point $\mathbf{p}$ to all other points $\mathbf{p}_1$, where p and p_1 both have magnitude k_F:

$$\int d^3q \to \frac{1}{v_F}\int d^2q \int_{-\infty}^{\infty} d\xi_1 \tag{8.223}$$

The limits on the $d\xi_1$ integral are extended between $\pm\infty$, since most of the integrand has large q values, where the actual limits on ξ_1 are from a very negative number to a very positive one. Since the main contribution is in the region $\xi_1 \approx 0$, this error is small. Consider the other angular factors in the integrand:

$$\frac{\mathbf{p}\cdot(\mathbf{p}+\mathbf{q})}{p^2} = 1 + \frac{q\nu}{p} = 1 + \frac{1}{2p^2}[p_1^2 - p^2 - q^2] \tag{8.224}$$

$$\approx 1 - \frac{q^2}{2k_F^2} + O\left(\frac{\xi_1}{E_F}, \frac{\xi_p}{E_F}\right) \tag{8.225}$$

The terms ξ_1/E_F, ξ_p/E_F are neglected, since it is expected that ξ_1 and ξ_p will be small—on the order of a Debye energy. On the other hand, the factor $q^2/2k_F^2$ need not be small, since the integration over phonon states has a significant contribution from high values of q near the edge of the Brillouin zone. This term must be retained. The difference in treating the factor $(p_1^2 - p^2)/k_F^2$, which is neglected, and q^2/k_F^2, which is retained, is that the former enters the average over electron states and the latter over phonon states. The integrals over the phonon wave vector can be expressed in terms of the function $\alpha^2F(\omega)$, which was introduced in Sec. 7.4:

$$\alpha^2F(\omega) = \frac{1}{v_F}\sum_\lambda \int \frac{d^3q}{(2\pi)^3} M_\lambda(\mathbf{q})^2 \delta[\omega - \omega_\lambda(\mathbf{q})] \tag{8.226}$$

$$\alpha_t^2F(\omega) = \frac{1}{v_F}\sum_\lambda \frac{d^3q}{(2\pi)^3} M_\lambda(\mathbf{q})^2 \frac{q^2}{2k_F^2} \delta[\omega - \omega_\lambda(\mathbf{q})] \tag{8.227}$$

The first of these is just the McMillan function, which was defined in Sec. 7.4. Since the Fermi surface has been assumed to be spherical, α^2F has the same value at each point on the surface, and the $\mathbf{p}$ subscript is omitted. The other form of coupling α_t^2F is called the "transport form of alpha-squared-F," which was introduced by Allen (1971). It differs from the McMillan form by having the additional factor of $q^2/2k_F^2$ in the integrand, which gives

more weight to the scattering processes at large wave vector. The factor of $q^2/2k_F^2$ is identical to $(1-\cos\theta')$ when the scattering is elastic.

After completing all these angular and wave vector integrations, the vertex function $\gamma(\mathbf{p}, ip, ip+i\omega)$ is found to be not very dependent on p. The only variation is as a function of p, which can be set equal to k_F with an error of only ξ_p/E_F. Define $\gamma(ip, ip+i\omega) = \gamma(k_F, ip, ip+i\omega)$ and arrive at the equations:

$$\gamma(ip, ip+i\omega) = 1 + \int_0^{\omega_D} du[\alpha^2F(u) - \alpha_t^2F(u)] \times \int_{-\infty}^{\infty} d\xi_1 S(\xi_1, u; ip, ip+i\omega) \tag{8.228}$$

$$S(\xi, u; ip, ip+i\omega) = \frac{1}{\beta}\sum_{iq}\frac{2u}{(iq)^2-u^2}\gamma(ip+iq, ip+iq+i\omega) \times \mathscr{G}(\xi, ip+iq)\mathscr{G}(\xi, ip+iq+i\omega) \tag{8.229}$$

This integral equation is not too difficult to solve. The basic approximation has been to decouple the integrations over dq and $d\xi_1$, which permits all the phonon information to be collected into the functions α^2F, α_t^2F. The primary assumption in this decoupling is that the Fermi degeneracy E_F is very much larger than other energies such as k_BT or $\hbar\omega_D$. The present integral equation is actually much easier to solve than the one for polarons in Sec. 8.2, since there the integration variables cannot be accurately decoupled in the same way.

The preceding is the basic integral equation which needs to be solved for the vertex function. The ξ dependence of S is only in the Green's functions $\mathscr{G}(\xi, ip+iq)$ and $\mathscr{G}(\xi, ip+iq+i\omega)$. This integral is done later, using the result (8.213) that only the integral over the pair $G_{\text{ret}}G_{\text{adv}} \equiv |G_{\text{ret}}|^2$ is nonzero. The factor $2u/[(iq)^2-u^2]$ is the phonon Green's function for a phonon of energy u. The next step in the derivation is to do the summation over Matsubara frequency iq, which is done in the usual way by constructing a contour integral.

$$\oint\frac{dz}{2\pi i}\frac{2u}{z^2-u^2}n_B(z)\gamma(z+ip, z+ip+i\omega)G(z+ip)G(z+ip+i\omega)$$

The contour in integration is a circle at infinity. The integrand has poles from the phonon Green's functions at $z=\pm u$ and cuts along the axes where the electron Green's functions are real, $z=-ip+\varepsilon'$ and $z=-ip-i\omega+\varepsilon'$. The contour integrals from these three contributions are

$$S = S_1 + S_2 + S_3 \tag{8.230}$$

$$S_1 = n_B(u)\gamma(ip+u, ip+i\omega+u)G(ip+u)G(ip+i\omega+u) + [n_B(u)+1]\gamma(ip-u, ip+i\omega-u)G(ip-u)G(ip+i\omega-u) \tag{8.231}$$

$$S_2 = \int\frac{d\varepsilon'}{2\pi i}\frac{2u}{(\varepsilon'-ip)^2-u^2}G(\varepsilon'+i\omega)[\gamma(\varepsilon'+i\delta, \varepsilon'+i\omega)G_{\text{ret}}(\varepsilon') - \gamma(\varepsilon'-i\delta, \varepsilon'+i\omega)G_{\text{adv}}(\varepsilon')] \tag{8.232}$$

$$S_2 = \int\frac{d\varepsilon'}{2\pi i}\frac{2u}{(\varepsilon'-ip-i\omega)^2-u^2}G(\varepsilon'-i\omega)[\gamma(\varepsilon'-i\omega, \varepsilon'+i\delta)G_{\text{ret}}(\varepsilon') - \gamma(\varepsilon'-i\omega, \varepsilon'-i\delta)G_{\text{adv}}(\varepsilon')] \tag{8.233}$$

The next step is to perform the analytic continuations. The final result should be $\Lambda(\varepsilon) = \gamma(\varepsilon - i\delta, \varepsilon + i\delta)$. First set $ip \to \varepsilon - i\delta$. The next step is to set $ip + i\omega \to \varepsilon + i\omega \to \varepsilon + \omega + i\delta \to \varepsilon + i\delta$, since $\omega \to 0$. The Green's functions become advanced or retarded according to the side of the cut, so this analytical continuation produces

$$\begin{aligned} S = {} & n_B(u)\Lambda(\varepsilon + u)|G_{\text{ret}}(\xi, \varepsilon + u)|^2 + [n_B(u) + 1]\Lambda(\varepsilon - u)|G_{\text{ret}}(\xi, \varepsilon - u)|^2 \\ & + \int \frac{d\varepsilon'}{2\pi i} n_F(\varepsilon')\left[|G_{\text{ret}}(\xi, \varepsilon')|^2 \Lambda(\varepsilon')\left(\frac{2u}{(\varepsilon' - \varepsilon - i\delta)^2 - u^2} - \frac{2u}{(\varepsilon' - \varepsilon + i\delta)^2 - u^2}\right)\right. \\ & \left. + O(G_{\text{ret}}^2, G_{\text{adv}}^2)\right] \end{aligned} \tag{8.234}$$

The next step is to do the integration over $d\xi_1$. According to (8.213), this integral eliminates all combinations of the Green's functions except $G_{\text{ret}}G_{\text{adv}} \equiv |G_{\text{ret}}|^2$, which removes the terms $O(G_{\text{ret}}^2, G_{\text{adv}}^2)$. Also note that the factor below is the phonon spectral function

$$\frac{2u}{(\varepsilon' - \varepsilon - i\delta)^2 - u^2} - \frac{2u}{(\varepsilon' - \varepsilon + i\delta)^2 - u^2} = 2\pi[\delta(\varepsilon' - \varepsilon - u) - \delta(\varepsilon' - \varepsilon + u)]$$

which eliminates the integral over $d\varepsilon'$. These manipulations provide the final form of the integral equation for $\Lambda(\varepsilon)$. It was first derived by Holstein (1964), although his result is modified by expressing it in the $\alpha^2 F$ formalism:

$$\begin{aligned} \Lambda(\varepsilon) = {} & 1 + \pi \int_0^{\omega_D} du[\alpha^2 F(u) - \alpha_t^2 F(u)] \\ & \times \left[[n_B(u) + n_F(\varepsilon + u)]\frac{\Lambda(\varepsilon + u)}{\Gamma(\varepsilon + u)} + [n_B(u) + 1 - n_F(\varepsilon - u)]\frac{\Lambda(\varepsilon - u)}{\Gamma(\varepsilon - u)}\right] \\ \sigma = {} & \frac{n_0 e^2}{2m}\int_{-\infty}^{\infty} d\varepsilon \frac{\Lambda(\varepsilon)}{\Gamma(\varepsilon)}\left[-\frac{dn_F(\varepsilon)}{d\varepsilon}\right] \end{aligned} \tag{8.235}$$

The integral (8.215) for the dc conductivity was rewritten in order to present the two important results together. The integral equation for the vertex function $\Lambda(\varepsilon)$ must be solved, and the solution is used in the integral over ε for the conductivity. Recall that $\Lambda(\varepsilon)$ is a real function, so the integral equation is not complicated. The form of the equations suggests the unknown quantity is actually $\bar{\tau}(\varepsilon) = 2\Lambda/\Gamma$, which might be called the effective relaxation time for transport.

The energy uncertainty $\Gamma(\varepsilon)$ of the electron from the electron–phonon interaction was given previously in (7.305). The equivalent result for the transport kernel is:

$$\begin{aligned} \Gamma(\varepsilon) &= -\text{Im}[\Sigma] = \pi \int_0^{\omega_D} du\alpha^2 F(u)[2n_B(u) + n_F(u + \varepsilon) + n_F(u - \varepsilon)] \\ \Gamma_t(\varepsilon) &= \pi \int_0^{\omega_D} du\alpha^2 F_t(u)[2n_B(u) + n_F(\varepsilon + u) + n_F(u - \varepsilon)] \end{aligned} \tag{8.236}$$

If the self-energy function Γ and the vertex function $\Lambda(\varepsilon)$ were both constants and independent of ε, the solution of the vertex equation (8.235) would be easy:

$$\Lambda = 1 + \frac{\Lambda}{\Gamma}[\Gamma - \Gamma_t] \tag{8.237}$$

$$= \frac{\Gamma}{\Gamma_t} \tag{8.238}$$

This model gives $\Lambda/\Gamma = 1/\Gamma_t$, so that the conductivity integral has only the transport form of the relaxation time. The earlier solution to impurity scattering also assumed that $\Lambda/\Gamma = 1/\Gamma_t$. For impurity scattering the functions $\Gamma(\varepsilon)$ and $\Lambda(\varepsilon)$ are usually insensitive to ε near $\varepsilon \approx 0$; the exception is when the impurity has a scattering resonance near the Fermi surface or if the density of states is not smooth. However, for the electron–phonon interaction in metals, it is not a good approximation to treat $\Gamma(\varepsilon)$, $\Gamma_t(\varepsilon)$ or $\Lambda(\varepsilon)$ as constants. The calculated results in Sec. 7.4, shown in Fig. 7.14 for Pb, illustrate that $\Gamma(\varepsilon)$ has substantial energy variations near the Fermi energy. The vertex function does also.

Takegahara and Wang (1977) evaluated (8.235) and (8.215) for metallic rubidium and cesium. Their results are shown in Fig. 8.11. In each case the solid line is calculated assuming that the ratio Γ/Γ_t is a constant, while the dashed line is calculated by solving (8.235) for the ε dependence of $\Lambda(\varepsilon)$. The latter curve is in very good agreement with the experiments, which are indicated by the points. The differences between the solid and dashed curves are similar to the differences between the Ziman formula (8.198) and the Holstein formula (8.235). Note that there is no region with a well-defined T^5 law for the resistivity in these metals.

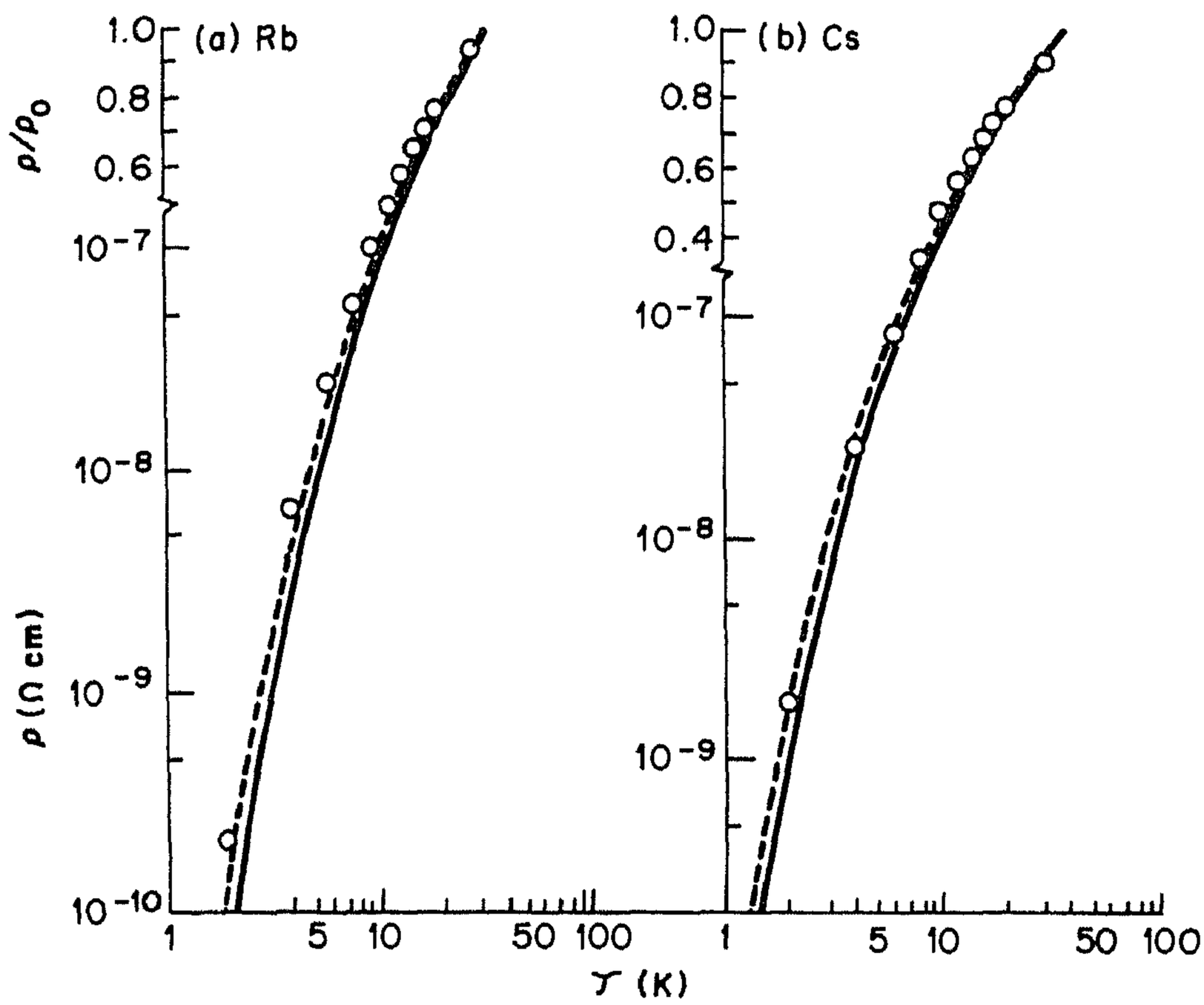


FIGURE 8.11 Calculated constant volume phonon limited electrical resistivity for (a) Rb and (b) Cs. Solid lines are calculations that treat $\Lambda(\varepsilon)$ as a constant. Dashed lines are full solution to integral equation for $\Lambda(\varepsilon)$. Points are experimental values. *Source*: Takegahara and Wang (1977) (used with permission).

8.4.3. Mass Enhancement

The electron–phonon mass enhancement factor λ was introduced in Sec. 7.4. It is from the real part of the electron self-energy due to the electron–phonon interaction. Since this self-energy was found to be energy dependent but not very wave vector dependent, the electron effective mass m^* is approximated by

$$\lambda(\omega) = -\frac{\partial}{\partial\omega}\Sigma(k_F, \omega) \tag{8.239}$$

$$\frac{m^*}{m} = 1 + \lambda(\omega) = \frac{1}{Z(\omega)} \tag{8.240}$$

The mass enhancement factor is also related to the quasiparticle renormalization factor $Z(\omega)$. The values of these quantities at the Fermi energy $\omega = 0$ are λ and Z_F. The mass renormalization factor λ can also be expressed as an average over the Fermi surface of a weighted average over the phonon density of states. This definition is given in Sec. 7.4. Values of λ in real metals range from 0.1 to 3.

An important question is the role which λ plays in the dc transport properties. The most obvious approximation is to use the effective mass $m^* = m(1 + \lambda)$ in transport formulas whenever classical theory says to use m. This substitution would make sense, since m^* is the effective mass which governs the motion of electrons on the Fermi surface, and these are involved in dc transport properties. However, this sensible procedure is wrong in most cases. The important point is that the factor $1 + \lambda$ can enter the final formulas in several ways. It changes the effective mass, relaxation time, and quasiparticle renormalization factor Z_F. The formula for the transport coefficient will have a number of factors of $1 + \lambda$. Often they all cancel, which is the case for the electrical conductivity.

Prange and Kadanoff (1964) investigated which transport measurements were influenced by the electron–phonon mass enhancement factor $1 + \lambda$. They concluded that the enhancement did affect the following measurements: specific heat, low-field cyclotron resonance, and the amplitude of the deHaas–van Alphen effect. The following measurements are not affected: dc electrical conductivity, thermoelectric power, thermal conductivity, the period of the deHaas–van Alphen effect, spin susceptibility, and the electron tunneling rate. Their conclusion on the thermoelectric power was challenged by Opsal *et al.* (1976), who detected a dependence on $1 + \lambda$. The list of quantities which are affected is much shorter than the list of quantities which are not affected. The usual case is that the transport property is not influenced by the mass enhancement factor; see Grimvall (1981) for a further discussion.

8.4.4. Thermoelectric Power

The thermoelectric power is a subject which is seldom discussed in most solid-state courses, as the lecturer is busy treating subjects which are more fashionable. This tendency is regrettable, since it is an important measurement. Experiments show that diverse behaviors are found for simple metals, and even the sign of this quantity shows no regularity. The low-temperature theory in metals is poorly understood, although there has been extensive work.

The thermoelectric power was introduced briefly in Sec. 3.8. It is a simple measurement, at least conceptually. A conducting bar is insulated so electrical currents cannot exit from its ends, and then a temperature difference ΔT is maintained along the length of the bar. The two

ends of the bar are found to have a voltage difference ΔV which is proportional to ΔT. The constant of proportionality is the thermoelectric power S:

$$S = -\frac{\Delta V}{\Delta T} = -\frac{1}{T}\frac{L^{(12)}}{L^{(11)}} \tag{8.241}$$

$$L^{(11)} = \frac{k_B T}{3i\omega\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \mathbf{j}(\tau)\cdot\mathbf{j}(0)\rangle \tag{8.242}$$

$$L^{(12)} = \frac{k_B T}{3i\omega\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \mathbf{j}_Q(\tau)\cdot\mathbf{j}(0)\rangle \tag{8.243}$$

The thermopower is defined theoretically as the ratio of two correlation functions. One is just proportional to the dc electrical conductivity $\sigma = \beta L^{(11)}$, which has already been evaluated for several models. The other correlation function has the argument of heat current $\mathbf{j}_Q$ and the electrical current $\mathbf{j}$, where $\mathbf{j}_Q = \mathbf{j}_E - (\mu/e)\mathbf{j}$. Equation (8.241) differs by a factor of charge e from its earlier version in Sec. 3.9, since earlier $\mathbf{j}$ meant the particle current, whereas now it is the electrical current. They differ by the unit of charge, which causes the change in (8.241).

The correlation function $L^{(12)}$ is now evaluated for its dc value. Both correlation functions are evaluated for values of $i\omega$; then analytically continue $i\omega \to \omega + i\delta$. Finally take the limit $\omega \to 0$ of the imaginary part of the retarded function.

In the evaluation of most correlation functions, there is usually a leading term which provides the dominant part of the answer. There are numerous small correction terms which can usually be ignored. In calculating the correlation function $L^{(12)}$ for the thermoelectric power, the dominant term vanishes, and one is left with obtaining all the numerous small correction terms. This feature makes it difficult to obtain an accurate answer.

The heat current operator has many terms. The one which is expected to provide the dominant term is from the kinetic energy of the electron: $\mathbf{j}_Q = \sum \mathbf{v}_\mathbf{p}\xi_\mathbf{p} C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}$. This heat current operator is used in most theories. Similarly, for the electrical current the operator is $\mathbf{j} = e\sum v_\mathbf{p} C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}$. The correlation function is called $L^{(12a)}$:

$$L^{(12a)} = \frac{1}{\nu\beta i\omega}\frac{e}{3m^2}\sum_{\mathbf{k}\mathbf{p}\sigma}\mathbf{k}\cdot\mathbf{p}\xi_\mathbf{p}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau C^\dagger_{\mathbf{k}\sigma}(\tau)C_{\mathbf{k}\sigma}(\tau)C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma}\rangle$$

The important feature of this correlation function is the bracket containing four electron operators. This type of operator sequence was encountered in earlier sections of this chapter. It is evaluated as

$$L^{(12a)} = \frac{1}{\nu\beta^2 i\omega}\frac{2e}{3m^2}\sum_\mathbf{k}\xi_\mathbf{k}\sum_{ik}\mathbf{k}\cdot\mathbf{\Gamma}(\mathbf{k}, ik, ik+i\omega)\mathcal{G}(k)\mathcal{G}(k+i\omega)$$

where $\mathbf{\Gamma}(\mathbf{k}, ik, ik+i\omega)$ is the vector vertex function of the bubble diagram. This same vertex function enters into the correlation function for the conductivity. Write it again as $\mathbf{\Gamma} = \mathbf{k}\gamma$. There is the same summation over Matsubara frequency ik which was done in the previous

section for the correlation function $L^{(11)}$. The result was given in (8.204). The retarded form of this correlation function is

$$\begin{aligned}\mathrm{Im}[L_{\mathrm{ret}}^{(12a)}] = -\frac{2e}{3\beta m^2\omega}\int\frac{d^3k}{(2\pi)^3}k^2\xi_{\mathbf{k}}\int_{-\infty}^{\infty}\frac{d\varepsilon}{2\pi}[n_F(\varepsilon+\omega)-n_F(\varepsilon)]\\ \times\{G_{\mathrm{adv}}(\mathbf{k},\varepsilon)G_{\mathrm{ret}}(\mathbf{k},\varepsilon+\omega)\gamma(\mathbf{k},\varepsilon-i\delta,\varepsilon+\omega+i\delta)\\ -\mathrm{Re}[G_{\mathrm{ret}}(\mathbf{k},\varepsilon)G_{\mathrm{ret}}(\mathbf{k},\varepsilon+\omega)\gamma(\mathbf{k},\varepsilon+i\delta,\varepsilon+\omega+i\delta)]\}\end{aligned}$$

The next step is to take the limit $\omega\to 0$, which causes the electron occupation factors to become $dn_F(\varepsilon)/d\varepsilon$. The integral over wave vector is changed to an integral over $k^2d^3k = 12\pi^3mn_0d\xi_k$:

$$\begin{aligned}\mathrm{Im}[L_{\mathrm{ret}}^{(12a)}] = \frac{en_0}{\beta m}\int_{-\infty}^{\infty}\frac{d\varepsilon}{2\pi}\left[-\frac{dn_F(\varepsilon)}{d\varepsilon}\right]\int_{-\infty}^{\infty}d\xi\xi\{|G_{\mathrm{ret}}(\mathbf{k},\varepsilon)|^2\\ \times\gamma(\mathbf{k},\varepsilon-i\delta,\varepsilon+i\delta)-\mathrm{Re}[G_{\mathrm{ret}}(\mathbf{k},\varepsilon)^2\gamma(\mathbf{k},\varepsilon+i\delta,\varepsilon+i\delta)]\}\end{aligned}\tag{8.244}$$

Again it is assumed that the retarded and advanced Green's functions are significantly dependent only on ξ in their kinetic energy term and that the self-energy and vertex functions have negligible dependence on ξ. Then the following integrals are evaluated by a contour integration in analogy with (8.213):

$$\begin{aligned}\int_{-\infty}^{\infty}d\xi\xi G_{\mathrm{ret}}(\xi,\varepsilon)^2 &= i\pi\\ \int_{-\infty}^{\infty}d\xi\xi|G_{\mathrm{ret}}(\xi,\varepsilon)|^2 &= \frac{\pi\Omega(\varepsilon)}{\Gamma(\varepsilon)}\\ \int_{-\infty}^{\infty}d\xi\xi G_{\mathrm{adv}}(\xi,\varepsilon)^2 &= -i\pi\end{aligned}\tag{8.245}$$

Two of the integrals equal $\pm i\pi$, which comes from the semicircle closing the contour at infinity. These integrals give zero when taking the real part. The important contribution must arise from the combination $|G_{\mathrm{ret}}(\xi,\varepsilon)|^2$, since it has a nontrivial contribution from the integral over $d\xi$:

$$\mathrm{Im}[L_{\mathrm{ret}}^{(12a)}] = \frac{en_0}{2\beta m}\int_{-\infty}^{\infty}d\varepsilon\left[-\frac{dn_F(\varepsilon)}{d\varepsilon}\right]\frac{\Omega(\varepsilon)\Lambda(\varepsilon)}{\Gamma(\varepsilon)} = 0\tag{8.246}$$

This integral is zero after one evaluates the integral $d\varepsilon$. It vanishes because the integrand is an antisymmetric function of ε. The quantities $dn_F(\varepsilon)/d\varepsilon$, $\Lambda(\varepsilon)$, and $\Gamma(\varepsilon)$ are all symmetric functions of ε, while $\Omega(\varepsilon)$ is an antisymmetric function. The correlation function $L^{(12a)}$ is zero.

The integral vanishes because of the single power of ξ in the $d\xi$ integral. It makes the single power of $\Omega(\varepsilon)$ in the integral over $d\varepsilon$. The single power of ξ comes from the heat current operator.

A nonzero result for the correlation function is obtained by repeating this derivation and retaining all the correction terms. One important term is from the argument of the wave vector integration:

$$\begin{aligned}k^2d^3k = 2\pi k^3dk^2 &= 3\pi m d\xi_k[k_F^2+(k^2-k_F^2)]^{3/2}\\ &\approx 4\pi k_F^3d\xi_k\left(1+\frac{3\xi k}{2E_F}\right)\end{aligned}\tag{8.247}$$

The second term in parentheses makes a nonzero contribution to the thermopower. This term is now evaluated. Kinetic energy integrals such as (8.245) must now be done, except there is a factor of ξ^2 in the integrand which multiplies the Green's functions. The evaluation of the integrand is tricky, since technically the integral diverges. At large values of ξ, then $G \sim 1/\xi$ and $(\xi G)^2 \sim 1$. The integrand does not fall off at large values of ξ, and taking the limits to $\pm\infty$ gives an infinite integral. Usually this problem is solved by ignoring it. The product $|G_{\text{ret}}|^2 = A/2\Gamma \approx \pi\delta(\xi - \varepsilon)/\Gamma$ is replaced by a delta function for energy conservation. The integral over $(\xi G_{\text{ret}})^2$ is set equal to zero:

$$\int d\xi \xi^2 |G_{\text{ret}}(\xi, \varepsilon)|^2 \approx \frac{\pi\varepsilon^2}{\Gamma(\varepsilon)} \tag{8.248}$$

The term in ε^2 is thermally averaged according to (8.244) using G&R 3.531(3);

$$\int_{-\infty}^{\infty} d\varepsilon \varepsilon^2 \left[-\frac{dn_F(\varepsilon)}{d\varepsilon} \right] = \frac{\pi^2}{3} \tag{8.249}$$

$$\text{Im}[L_{\text{ret}}^{(12a)}] = \frac{\pi^2 e n_0 \tau_t}{3\beta m E_F} \eta(T) \tag{8.250}$$

$$S = -\left(\frac{k_B}{\sigma}\right) \text{Im}[L_{\text{ret}}^{(12a)}] = -\frac{\pi^2 k_B^2 T}{3eE_F} \eta(T) \tag{8.251}$$

where the conductivity is $\sigma = n_0 e^2 \tau_t / m$. The dimensions of S are volts per degree. The parameter $\eta(T)$ is dimensionless. So far $\eta = \frac{3}{2}$ from the coefficient of ξ/E_F in (8.247). Other contributions to $\eta(T)$ are derived below.

Taylor and MacDonald (1986) evaluated this expression for the alkali metals at high temperature. Rather good agreement is obtained, as shown below. First it is necessary to find more contributions to $\eta(T)$. At high temperature it is a good approximation to set $\tau_t = \tau_{pt}$, where the transport lifetime is entirely from phonons, and is given by the transport form of alpha-squared-f in (8.226):

$$\frac{1}{\tau_{pt}(k)} = \frac{m}{(2\pi)^2 k^3} \int_0^{2k} q^3 dq \int_0^{2\pi} d\phi \sum_\lambda M_\lambda^2(\mathbf{q}) \{ n_B[\omega_\lambda(\mathbf{q})] + n_F[\omega_\lambda(\mathbf{q})] \}$$

Contributions to $\eta(T)$ are obtained by expanding k about k_F and keeping the first-order terms in $(k - k_F)/k_F \approx \xi/2E_F$. The prefactor of k^{-3} gives another contribution of $\eta = \frac{3}{2}$ to η. The integration limit of $2k$ gives a contribution to η of

$$2q(T) = \frac{2m}{\pi^2} \tau_{pt} \int_0^{2\pi} d\phi \sum_\lambda M_\lambda^2(2k_F) \{ n_B[\omega_\lambda(2k_F)] + n_F[\omega_\lambda(2k_F)] \}$$

Another contribution to $\eta(T)$ comes from the matrix element $M_\lambda(\mathbf{q})$. It is usually calculated using a screened pseudopotential for the electron–ion interaction. The better pseudopotentials are nonlocal, which means they depend upon $M_\lambda(\mathbf{k}, \mathbf{k} + \mathbf{q})$ rather than just on $\mathbf{q}$. This k dependence can also be expanded around the point k_F. The pseudopotential gives another contribution to $\eta(T)$ which is called $r(T)$. The contributions to $\eta(T)$ are

$$\eta(T) = 3 - 2q(T) - \tfrac{1}{2} r(T) \tag{8.252}$$

Table 8.2 shows the evaluation of these terms for the alkali metals at various temperatures. The comparison with the experimental data is good for Na, K, and Rb. The high-temperature thermopower seems to be understood in these cases. Both $q(T)$ and $r(T)$ are small for Na

TABLE 8.2 Thermoelectric parameters in the alkali metals. $\eta = 3 - 2q - r/2$ (Taylor and MacDonald, 1986, used with permission).

Metal	$T(K)$	$2q(T)$	$r(T)/2$	η	
				Theory	Experiment
Li	424	9.26	−1.43	−5.33	−6.3
Na	300	0.04	−0.09	3.05	2.9
K	200	0.83	−1.87	4.04	4.0
Rb	100	4.78	−4.49	2.71	2.8
Cs	100	9.32	−7.15	0.83	0.0

because the electron–ion pseudopotential is nearly zero at $q = 2k_F$. The thermopower is easier to evaluate at high rather than at low temperature. At high temperature the resistance is dominated by the electron–phonon interaction, which is well approximated by using the transport form of alpha-squared-f. At low temperatures one also has to include the k dependence of impurity scattering, as well as the ordinary form of alpha-squared-f while solving (8.235). In addition, phonon drag is important at low temperature.

8.5. QUANTUM BOLTZMANN EQUATION

There are several different methods of doing transport theory. The theory used in the preceding sections uses the Kubo relation for the conductivity and is called "linear response." One assumes that currents are proportional to fields. The proportionality constants can be evaluated in equilibrium. This method works because one assumes that the applied fields are small, and the system is only infinitesimally disturbed from equilibrium.

A second method of transport theory is discussed in this section. One assumes the existence of a distribution function f, which describes the behavior of the particles. One writes a differential equation for the motion of f through phase space. The differential equation is a Boltzmann equation. One then tries to solve the Boltzmann equation for a system out of equilibrium. For fields that are small, the system is only slightly out of equilibrium, and one reproduces the linear response solutions described earlier. The advantage of the Boltzmann equation method is that one can also try to solve the equation when the system is far from equilibrium.

The original Boltzmann equation described the behavior of a distribution function $f(\mathbf{v}, \mathbf{R}, t)$ of three variables: velocity, position, and time. The Wigner distribution function (WDF) was introduced in Sec. 3.7. It is equivalent to a distribution function $f(\mathbf{k}, \omega; \mathbf{R}, t)$ with four variables: wave vector $\mathbf{k}$, energy ω, position $\mathbf{R}$, and time t. This latter distribution function is the one needed for many-particle systems. Since $f(\mathbf{k}, \omega; \mathbf{R}, t)$ is not positive definite, calling it a "distribution function" is probably misleading. This phrase is used since it is widespread, but the warning should be kept in mind.

The differential equation obeyed by $f(\mathbf{k}, \omega; \mathbf{R}, t)$ is called the *quantum Boltzmann equation*, which is abreviated QBE. It is derived rigorously in the following sections. Here a

quick derivation is provided using semiclassical arguments. According to the Liouville theorem a distribution function $f(q_i, t)$ is stationary when it obeys the equation

$$\delta f = 0 = \frac{\partial f}{\partial t} + \sum_i \frac{\partial f}{\partial q_i}\dot{q}_i + \left(\frac{\partial f}{\partial t}\right)_s \tag{8.253}$$

where the last term is from scattering. For the WDF, $\dot{\mathbf{k}} = \mathbf{F}$ (force), $\dot{\mathbf{R}} = \mathbf{v}$ (velocity), and $\dot{\omega} = \mathbf{v}\cdot\mathbf{F}$ (Joule heating). The QBE is

$$0 = \left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla_R + \mathbf{F}\cdot\nabla_k + \mathbf{v}\cdot\mathbf{F}\frac{\partial}{\partial\omega}\right)f + \left(\frac{\partial f}{\partial t}\right)_S \tag{8.254}$$

This equation has nearly the right form. It is derived rigorously in the next section, which produces a few more terms from self-energy contributions. That derivation also provides a prescription for obtaining the scattering term.

Equation (8.254) has one feature that is important. The additional variable ω also causes a new driving term on the left of the form $\mathbf{v}\cdot\mathbf{F}\partial/\partial\omega$. This term was first derived by Mahan and Hansch (1983). The semiclassical distribution function $f(\mathbf{v}, \mathbf{R}, t)$ lacks this driving term since it lacks the energy variable ω. The QBE for the WDF is a different equation from the traditional Boltzmann equation.

8.5.1. Derivation of the QBE

The QBE is the equation of motion for the Green's function $G^<$. The method of deriving transport equations was pioneered by Kadanoff and Baym (1962). Recall from Sec. 3.7 that $f(\mathbf{r}, t; \mathbf{R}, T) = -iG^<(\mathbf{r}, t; \mathbf{R}, T)$, where $(\mathbf{r}, t)$ are the relative variables and $(\mathbf{R}, T)$ are the position and time in center-of-mass. The Green's function was defined in terms of the field operator as

$$G^<(\mathbf{r}, t; \mathbf{R}, T) = i\langle\psi^\dagger(\mathbf{R} - \tfrac{1}{2}\mathbf{r}, T - \tfrac{1}{2}t)\psi(\mathbf{R} + \tfrac{1}{2}\mathbf{r}, T + \tfrac{1}{2}t)\rangle \tag{8.255}$$

The next step is to Fourier transform the relative variables $(\mathbf{r}, t)$ into $(\mathbf{k}, \omega)$:

$$G^<(\mathbf{k}, \omega; \mathbf{R}, T) = \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \int dt e^{i\omega t} G^<(\mathbf{r}, t; \mathbf{R}, T) \tag{8.256}$$

The QBE will be derived for a particle in a weak electric field. The intent is to describe interacting many-particle systems that have a small current flowing in response to a small electric field. The derivation will be sufficiently general to include any kind of particles and nearly any kind of interactions.

The electric field can be introduced as either a scalar or a vector potential. The QBE is independent of this choice, as required by gauge invariance. Here both are included, in order to provide the most general derivation. There will be an electric field $\mathbf{E}_v$, which is from a vector potential, and another electric field $\mathbf{E}_s$ from a scalar potential. The final version of the QBE will include only the total electric field $\mathbf{E} = \mathbf{E}_v + \mathbf{E}_s$. The scalar potential is introduced through the interaction term

$$H_E = -e\mathbf{E}_s \cdot \sum_j \mathbf{r}_j \tag{8.257}$$

The vector potential is introduced by changing the momentum of each charged particle to $(\mathbf{p} - e\mathbf{A}/c)$, where the vector potential is $\mathbf{A} = -c\mathbf{E}_v t$. A vector potential proportional to time

could occur in a wire loop with a slowly varying magnetic flux through the center. The present theory is for a dc electric field.

Equations (3.331) and (3.332) are equations of motion for the four Green's functions in the 2×2 matrix for $\tilde{G}$. The two electric field terms are added to the left-hand side of these equations—they are included in H_0. The derivation of the QBE involves several algebraic manipulations on the left-hand side of these equations. In order to avoid a lot of cumbersome notation, the right-hand side of these equations are not written during these steps. On the left, the steps are the same for all six of the real-time Green's functions. The generic symbol G applies to any one of them.

The two equations for G in (3.331) and (3.332) are first added and then subtracted:

$$\left[i\left(\frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2}\right) - H_1 - H_2\right]G = \tag{8.258}$$

$$\left[i\left(\frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_2}\right) - H_1 + H_2\right]G = \tag{8.259}$$

where $H_1 = H_0(\mathbf{r}_1, \mathbf{p}_1)$ and $H_2 = H_0(\mathbf{r}_2, -\mathbf{p}_2)$. The two equations contain time derivatives that relate either to the relative or center-of-mass motion

$$\frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_2} = \frac{\partial}{\partial T}, \qquad \frac{\partial}{\partial t_1} - \frac{\partial}{\partial t_2} = 2\frac{\partial}{\partial t} \tag{8.260}$$

For particles with parabolic band dispersion, the sum and difference of the two Hamiltonians produce simple expressions in relative coordinates. As a first step, consider what happens to the two momentum terms which are in the form of $(\mathbf{p} - e\mathbf{A}/c)$

$$\mathbf{p}_1 + e\mathbf{E}_v t_1 = (\mathbf{p} + e\mathbf{E}_v T) + \tfrac{1}{2}(\mathbf{P} + e\mathbf{E}_v t) \tag{8.261}$$

$$\mathbf{p}_2 - e\mathbf{E}_v t_2 = -(\mathbf{p} + e\mathbf{E}_v T) + \tfrac{1}{2}(\mathbf{P} + e\mathbf{E}_v t) \tag{8.262}$$

where $\mathbf{p}$ and $\mathbf{P}$ are the relative and center-of-mass momentum. It is important to understand all of the various plus and minus signs. Since $\mathbf{A} = -c\mathbf{E}_v t$ then $(\mathbf{p} - e\mathbf{A}/c) = (\mathbf{p} + e\mathbf{E}_v t)$ for $\mathbf{p}_1$. However, in H_2 the momentum enters as $-\mathbf{p}_2$ which explains the sign change on the bottom. On the right one uses $t_{1,2} = T \pm t/2$ and $\mathbf{p}_{1,2} = \pm\mathbf{p} + \mathbf{P}/2$. These results make it easy to see the form of $H_1 \pm H_2$:

$$H_1 + H_2 = \frac{1}{m}(\mathbf{p} + e\mathbf{E}_v T)^2 + \frac{1}{4m}(\mathbf{P} + e\mathbf{E}_v t)^2 - 2e\mathbf{E}_s \cdot \mathbf{R} \tag{8.263}$$

$$H_1 - H_2 = \frac{1}{m}(\mathbf{p} + e\mathbf{E}_v T) \cdot (\mathbf{P} + e\mathbf{E}_v t) - e\mathbf{E}_s \cdot \mathbf{r} \tag{8.264}$$

Divide the top equation by two, and obtain the following two equations for the Green's function:

$$2\left[i\frac{\partial}{\partial t} - \frac{1}{2m}(\mathbf{p} + e\mathbf{E}_v T)^2 - \frac{1}{m}(\mathbf{P} + e\mathbf{E}_v t)^2 + e\mathbf{E}_s \cdot \mathbf{R}\right]G = \tag{8.265}$$

$$\left[i\frac{\partial}{\partial T} - \frac{1}{m}(\mathbf{p} + e\mathbf{E}_v T) \cdot (\mathbf{P} + e\mathbf{E}_v t) + e\mathbf{E}_s \cdot \mathbf{r}\right]G = \tag{8.266}$$

These two equations describe the relative and center-of-mass motion of the function $G(\mathbf{r}, t; \mathbf{R}, T)$. The goal is to derive the QBE for the WDF $G^<(\mathbf{k}, \omega; \mathbf{R}, T)$. Fourier transform

the variables $(\mathbf{r}, t)$ to the set $(\mathbf{q}, \Omega)$ as in (8.256). This transform changes $\mathbf{p}$ to $\mathbf{q}$, $\mathbf{r}$ to $i\nabla_{\mathbf{q}}$, d/dt to $-i\Omega$, and t to $-i\partial/\partial\Omega$. The two transformed equations are

$$\begin{aligned} &2\left[\Omega + e\mathbf{E}_s\cdot\mathbf{R} - \frac{1}{2m}(\mathbf{q}+e\mathbf{E}_v T)^2 + \frac{1}{8m}\left(\nabla_R + e\mathbf{E}_v\frac{\partial}{\partial\Omega}\right)^2\right]G(\mathbf{q},\Omega;\mathbf{R},T) = \\ &i\left[\frac{\partial}{\partial T} + \frac{1}{m}(\mathbf{q}+e\mathbf{E}_v T)\cdot\left(\nabla_R + \varepsilon\mathbf{E}_v\frac{\partial}{\partial\Omega}\right) + e\mathbf{E}_s\cdot\nabla_{\mathbf{q}}\right]G(\mathbf{q},\Omega;\mathbf{R},T) = \end{aligned} \tag{8.267}$$

The second of these equations has, on the left-hand side, exactly the same terms as are found in the Boltzmann equation. This similarity suggests that the QBE is the same as the BE. However, this conclusion is incorrect. The above set of equations have several things wrong with them

1. There is the term $e\mathbf{E}_s\cdot\mathbf{R}$. This term seems to combine with Ω to produce a center-of mass energy $\omega = \Omega + e\mathbf{E}_s\cdot\mathbf{R}$. Together they suggest that the energy of a particle depends upon its location. The energy is different at one end of a sample than at the other. However, this behavior is contrary to common sense. When there is a small electric field along the sample, and a small current flowing, we expect the system to be uniform. There is the same particle density, current density, etc. at each point in the solid. There is no dependence upon $\mathbf{R}$. This undesirable term has to be eliminated.
2. The relative momentum seems to enter in the combination of $\mathbf{q} + e\mathbf{E}_v T$. It depends upon the center-of-mass time T. This feature is also unphysical, and needs to be eliminated.
3. The result is not gauge invariant, since the two fields $\mathbf{E}_s$ and $\mathbf{E}_v$ enter differently.

All of these problems can be eliminated through a variable transformation. Of course, this transformation also causes some derivatives to change:

$$\begin{aligned} \Omega + e\mathbf{E}_s\cdot\mathbf{R} &\Rightarrow \omega \\ \mathbf{q} + e\mathbf{E}_v T &\Rightarrow \mathbf{k} \\ \nabla_R &\Rightarrow \nabla_R + e\mathbf{E}_s\frac{\partial}{\partial\omega} \\ \frac{\partial}{\partial T} &\Rightarrow \frac{\partial}{\partial T} + e\mathbf{E}_v\cdot\nabla_{\mathbf{k}} \end{aligned} \tag{8.268}$$

These transformations cause the two equations in (8.267) for the Green's function to now have the form

$$\begin{aligned} &\left[\omega - \varepsilon_{\mathbf{k}} + \frac{1}{8m}\left(\nabla_R + e\mathbf{E}\frac{\partial}{\partial\omega}\right)^2\right]G(\mathbf{k},\omega;\mathbf{R},T) = \\ &i\left[\frac{\partial}{\partial T} + \mathbf{v}_{\mathbf{k}}\cdot\nabla_R + e\mathbf{E}\cdot\left(\nabla_{\mathbf{k}} + \mathbf{v}_{\mathbf{k}}\frac{\partial}{\partial\omega}\right)\right]\mathbf{G}(\mathbf{k},\omega;\mathbf{R},T) = \end{aligned} \tag{8.269}$$

The notation on the Green's function has been changed again. The arguments $(\mathbf{q}, \Omega)$ have been changed to $(\mathbf{k}, \omega)$. The left-hand sides of these two equations are now in the form that is useful. The lower equation has exactly the same terms as in (8.254), and is the quantum

Boltzmann equation. The electric field is $\mathbf{E} = \mathbf{E}_v + \mathbf{E}_s$. Both electric fields contribute in the same way, and the result is now gauge invariant.

The above variable transformation makes the QBE gauge invariant. The results are valid for dc electric fields. A more complicated variable transformation is required for ac electric fields (Levanda and Fleurov, 1994).

Now it is time to restore the scattering terms to the right-hand side of the equal sign. The scattering terms for $G^<$ and G_{ret} are in (3.341). They are the most important Green's functions in applications using real time. The self-energy terms on the right are found by following the same steps to bring the left-hand side to (8.269). In doing these operations, keep in mind that the scattering terms are of the form

$$\int dx_3 \Sigma(x_1, x_3) G(x_3, x_2)$$

First add and subtract the two equations for each Green's function. There follows a series of variable transformations. They can be understood by examining just one term in the right-hand side of Eq. (8.269). The various variable arguments yield a fairly complicated expression, which is derived by the following steps:

1. Change the two sets of variables $(x_1, x_3)(x_3, x_2)$ to the center-of-mass grouping $(x_1 - x_3, (x_1 + x_3)/2)(x_3 - x_2, (x_3 + x_2)/2)$.
2. Change the integration variable from x_3 to $y = x_1 - x_3$, so the variable grouping become $(y, x_1 - y/2)(x_1 - x_2 - y, (x_1 + x_2 - y)/2)$.
3. Change to center of mass variables $x = (x_1 - x_2), X = (x_1 + x_2)/2$, which produces the variable grouping $(y, X + (x - y)/2)(x - y, X - y/2)$.
4. Change the integration variable x to $z = x - y$ which produces the final arguments of $(y, X + z/2)(z, X - y/2)$.

Below are the full scattering equations for $G^<$ and G_{ret}. The explicit variables in the scattering terms are only written out in the first term, but are identical for the other three. The order of the terms in the scattering integral is important. The first one always has argument $(y, X + z/2)$ while the second one has argument $(z, X - y/2)$. The four-vector in the exponent is $q = (\mathbf{k}, \omega)$:

$$\begin{aligned}
&\left[\omega - \varepsilon_{\mathbf{k}} + \frac{1}{8m}\left(\nabla_R + e\mathbf{E}\frac{\partial}{\partial\omega}\right)^2\right] G_{\text{ret}}(\mathbf{k}, \omega; \mathbf{R}, T) \\
&\quad = 1 + \frac{1}{2}\int dz e^{-iqz} \int dy e^{-iqy} [\Sigma_{\text{ret}}(y, X + z/2) G_{\text{ret}}(z, X - y/2) \\
&\qquad + G_{\text{ret}}(y, X + z/2)\Sigma_{\text{ret}}(z, X - y/2)] \qquad (8.270)
\end{aligned}$$

$$\begin{aligned}
&i\left[\frac{\partial}{\partial T} + \mathbf{v}_{\mathbf{k}} \cdot \nabla_R + e\mathbf{E} \cdot \left(\nabla_{\mathbf{k}} + \mathbf{v}_{\mathbf{k}}\frac{\partial}{\partial\omega}\right)\right] G_{\text{ret}}(\mathbf{k}, \omega; \mathbf{R}, T) \\
&\quad = \int dz e^{-iqz} \int dy e^{-iqy} [\Sigma_{\text{ret}} G_{\text{ret}} - G_{\text{ret}}\Sigma_{\text{ret}}] \qquad (8.271)
\end{aligned}$$

$$\left[\omega - \varepsilon_{\mathbf{k}} + \frac{1}{8m}\left(\nabla_R + e\mathbf{E}\frac{\partial}{\partial\omega}\right)^2\right]G^<(\mathbf{k}, \omega; \mathbf{R}, T)$$

$$= \frac{1}{2}\int dz e^{-iqz}\int dy e^{-iqy}[\Sigma_t G^< - \Sigma^< G_{\bar{t}} + G_t\Sigma^< - G^<\Sigma_{\bar{t}}] \tag{8.272}$$

$$i\left[\frac{\partial}{\partial} + \mathbf{v}_{\mathbf{k}}\cdot\nabla_R + e\mathbf{E}\cdot\left(\nabla_{\mathbf{k}} + \mathbf{v}_{\mathbf{k}}\frac{\partial}{\partial\omega}\right)\right]G^<(\mathbf{k}, \omega; \mathbf{R}, T)$$

$$= \int dz e^{-iqz}\int dy e^{-iqy}[\Sigma_t G^< - \Sigma^< G_{\bar{t}} - G_t\Sigma^< + G^<\Sigma_{\bar{t}}] \tag{8.273}$$

These four equations are the important ones for nonequilibrium calculations. Although one primarily wants to find $G^<$, it is always necessary to first find G_{ret}. These equations were first derived by Fleurov and Kozlov (1978).

So far no approximations have been made, and the equations are exact. The QBE has a linear term in the electric field **E**. However, the equation is exact to all powers of **E**, not just to the first power.

8.5.2. Gradient Expansion

Equations (8.271) are usually too hard to solve because the scattering terms on the right have a complicated form. Some sort of approximation has to be introduced to simplify the right side. The approximation described below is only valid to first order in the field. Kadanoff and Baym (1962) introduced an approximation for evaluating these scattering terms, which is called the "gradient expansion." They assume that the center-of-mass time T is very large, and take the limit that $T \to \infty$. At large values of T, they assume that the system is approaching its asymptotic limit, so that variations with respect to T are small. Obviously the gradient expansion is not suitable for studying transients, since it is poor at small values of T. Neither is it useful for steady-state ac phenomena (Mahan, 1987). Indeed, the T dependence is so poorly described in the gradient approximation that it should not be used and the T derivative terms should be dropped from the QBE. Nevertheless, the gradient expansion is used since it is applicable for homogeneous (small R dependence) steady state (small T dependence) systems.

The center-of-mass variables are all in the form $(\mathbf{R} + \Delta\mathbf{R}, T + \Delta T)$, which are expanded in a Taylor series about the point $X \equiv (\mathbf{R}, T)$. The integrals can be done for each term in the series. These integrals usually cause further derivatives. The gradient expansion is shown below for a typical scattering term:

$$I = \int dz e^{-iqz}\int dy e^{-iqy}\Sigma(y, X + z/2)G(z, X - y/2) \tag{8.274}$$

$$= \int dz e^{-iqz}\int dy e^{-iqy}\left\{\Sigma(y, X)G(z, X)\right.$$

$$\left. + \frac{z}{2}\frac{\partial\Sigma(y, X)}{\partial X}G(z, X) - \Sigma(y, X)\frac{y}{2}\frac{\partial G(z, X)}{\partial X} + \cdots\right\} \tag{8.275}$$

$$= \Sigma(q, X)G(q, X) + \frac{i}{2}[(\nabla_q\Sigma)\nabla_X G - (\nabla_X\Sigma)\nabla_q G] + \cdots \tag{8.276}$$

The first term on the right are usually the largest. It is customary to retain the first derivative terms and ignore higher derivatives. They can be expressed using a Poisson bracket notation:

$$[\Sigma, G] = (\nabla_q \Sigma)\nabla_X G - (\nabla_X \Sigma)\nabla_q G \tag{8.277}$$

$$= \frac{\partial \Sigma}{\partial \Omega}\frac{\partial G}{\partial T} - \frac{\partial G}{\partial \Omega}\frac{\partial \Sigma}{\partial T} - \nabla_{\mathbf{q}}\Sigma \cdot \nabla_{\mathbf{R}} G + \nabla_{\mathbf{q}} G \cdot \nabla_{\mathbf{R}}\Sigma \tag{8.278}$$

These frequency derivatives are with respect to the Ω variables, since the variable change ($\mathbf{k} = \mathbf{q} + e\mathbf{E}_v T$, $\omega = \Omega + e\mathbf{E}_s \cdot \mathbf{R}$) has not yet been made. If this step is taken now, the derivatives get altered according to Eq. (8.268), which changes the Poisson brackets to

$$[C, D] \Rightarrow [C, D] + e\mathbf{E} \cdot \left[\left(\frac{\partial C}{\partial \omega}\right)\nabla_{\mathbf{k}} D - \left(\frac{\partial D}{\partial \omega}\right)\nabla_{\mathbf{k}} C\right] \tag{8.279}$$

where $\mathbf{E}$ is again the total electric field. This analysis finally derives from (8.271) the following expression for the nonequilibrium retarded Green's function in an electric field, when G_{ret} and Σ_{ret} depend upon ($\mathbf{k}$, ω; $\mathbf{R}$, T):

$$\left[\omega - \varepsilon_{\mathbf{k}} + \frac{1}{8m}\left(\nabla_R + e\mathbf{E}\frac{\partial}{\partial \omega}\right)^2 - \Sigma_{\text{ret}}\right] G_{\text{ret}}(\mathbf{k}, \omega; \mathbf{R}, T) = 1 \tag{8.280}$$

$$i\left\{\frac{\partial}{\partial T} + \mathbf{v}_{\mathbf{k}} \cdot \nabla_R + e\mathbf{E} \cdot \left[\left(1 - \frac{\partial \Sigma_{\text{ret}}}{\partial \omega}\right)\nabla_{\mathbf{k}} + (\mathbf{v}_{\mathbf{k}} + \nabla_{\mathbf{k}}\Sigma_{\text{ret}})\frac{\partial}{\partial \omega}\right]\right\} G_{\text{ret}} = i[\Sigma_{\text{ret}}, G_{\text{ret}}], \tag{8.281}$$

The additional terms from the Poisson bracket, which are linear in the field $\mathbf{E}$, have been transferred to the left of the equal sign.

These equations simplify for nonequilibrium systems which are both homogeneous ($\nabla_{\mathbf{R}} = 0$) and steady state ($\partial/\partial T = 0$). The Poisson brackets vanish, as do several terms on the left. Also ignored are terms nonlinear in the electric field $O(\mathbf{E}^2)$, and find for the above two equations

$$[\omega - \varepsilon_{\mathbf{k}} - \Sigma_{\text{ret}}] G_{\text{ret}} = 1 \tag{8.282}$$

$$ie\mathbf{E} \cdot \left[\left(1 - \frac{\partial \Sigma_{\text{ret}}}{\partial \omega}\right)\nabla_{\mathbf{k}} + (\mathbf{v}_{\mathbf{k}} + \nabla_{\mathbf{k}}\Sigma_{\text{ret}})\frac{\partial}{\partial \omega}\right] G_{\text{ret}} = 0 \tag{8.283}$$

The first equation is easily solved, to yield

$$G_{\text{ret}}(\mathbf{k}, \omega) = \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma_{\text{ret}}(\mathbf{k}, \omega)} + O(\mathbf{E}^2) \tag{8.284}$$

The retarded Green's function appears to have no first-order term in the electric field. It actually does, since the self-energy Σ_{ret} has a term linear in the field due to the electron–phonon interaction. This term is small and seems to have little effect. If it is ignored then the retarded Green's function is unchanged from its value in equilibrium. This result considerably simplifies the solution to the QBE. The solution (8.284) also satisfies the equation (8.283). Related quantities such as the advanced function $G_{\text{adv}} = G^*_{\text{ret}}$ and the spectral function $A(\mathbf{k}, \omega)$ are also unchanged to first order in the electric field. This completes the discussion of the retarded Green's function in a static electric field.

Next the gradiant expansion is applied to (8.271) for the equations for $G^<$. Again keeping only first-order derivatives, the final results are the quantum Boltzmann equation (QBE):

$$\left[\omega - \varepsilon_{\mathbf{k}} + \frac{1}{8m}\left(\nabla_R + e\mathbf{E}\frac{\partial}{\partial\omega}\right)^2\right]G^< = G^< \operatorname{Re}[\Sigma_{\text{ret}}] + \Sigma_{\text{ret}} \operatorname{Re}[G_{\text{ret}}]$$
$$+\frac{i}{4}[\Sigma^>, G^<] - \frac{i}{4}[\Sigma^<, G^>] + \frac{i}{4}e\mathbf{E}\cdot\left[\left(\frac{\partial\Sigma^>}{\partial\omega} - \frac{\partial\Sigma^<}{\partial\omega}\right)\nabla_{\mathbf{k}}G^<\right.$$
$$\left.+\left(\frac{\partial G^>}{\partial\omega} - \frac{\partial G^<}{\partial\omega}\right)\nabla_{\mathbf{k}}\Sigma^<\right] \tag{8.285}$$

$$i\left\{\frac{\partial}{\partial T} + \mathbf{v}_{\mathbf{k}}\cdot\nabla_R + e\mathbf{E}\cdot\left[\left(1 - \frac{\partial\Sigma_{\text{ret}}}{\partial\omega}\right)\nabla_{\mathbf{k}} + (\mathbf{v}_{\mathbf{k}} + \nabla_{\mathbf{k}}\Sigma_{\text{ret}})\frac{\partial}{\partial\omega}\right]\right\}G^<$$
$$- ie\mathbf{E}\cdot\left[\frac{\partial\Sigma^<}{\partial\omega}\nabla_{\mathbf{k}} \operatorname{Re}[G_{\text{ret}}] - \frac{\partial G_{\text{ret}}}{\partial\omega}\nabla_{\mathbf{k}}\Sigma^<\right] = \Sigma^> G^< - \Sigma^< G^>$$
$$+ i[\operatorname{Re}[\Sigma_{\text{ret}}], G^<] + i[\Sigma^<, \operatorname{Re}[G_{\text{ret}}]] \tag{8.286}$$

Standard relationships have been used in deriving this equation, such as $G_t - G_{\bar{t}} = 2\operatorname{Re}[G_{\text{ret}}]$ and $G_t + G_{\bar{t}} = G^< + G^>$.

Equation (8.286) is the quantum Boltzmann equation. It is rather formidable. It is also difficult to solve, since it is usually an integral equation. Sometimes it is nonlinear in the particle density because the self-energy functions $\Sigma^<$ and $\Sigma^>$ are also functions of $G^<$ and $G^>$.

The QBE also contains the functions $G^>$ and $\Sigma^>$. Similar equations for these functions can be derived by starting from the general equations (3.331) and (3.332). This derivation shows that the equation for $G^>$ is almost identical to the one for $G^<$. In fact, one can prove that the following identities are valid:

$$G^> = G^< - iA \tag{8.287}$$

$$\Sigma^> = \Sigma^< + 2i \operatorname{Im}[\Sigma_{\text{ret}}] = \Sigma^< - 2i\Gamma \tag{8.288}$$

These relations are trivial to show for equilibrium, but they are also valid for nonequilibrium situations. These identities will be used often to simplify expressions. For example, the main scattering term in the QBE can be immediately simplified to

$$\Sigma^> G^< - \Sigma^< G^> = -i\{2\Gamma G^< - \Sigma^< A\} \tag{8.289}$$

This result will be employed in the calculations. The quantities $G^<$ and $\Sigma^<$ are generally proportional to the density of particles, while retarded functions are only indirectly dependent upon the density of particles—only through the self-energy function Σ_{ret}. In the QBE, each term has one factor that is either $G^<$ or $\Sigma^<$, so each term is proportional to the density of particles. This equation does have the character of a transport equation.

The QBE simplifies for the treatment of systems that are homogeneous ($\nabla_R = 0$) and steady ($\partial/\partial T = 0$). These derivatives are dropped, as well as the Poisson brackets, since the latter contain similar derivatives. In this case the QBE is

$$e\mathbf{E}\cdot\left[\left(1-\frac{\partial\Sigma_{\text{ret}}}{\partial\omega}\right)\nabla_{\mathbf{k}}+(\mathbf{v}_{\mathbf{k}}+\nabla_{\mathbf{k}}\Sigma_{\text{ret}})\frac{\partial}{\partial\omega}\right]G^{<}$$
$$-ie\mathbf{E}\cdot\left[\frac{\partial\Sigma^{<}}{\partial\omega}\nabla_{\mathbf{k}}\,\text{Re}[G_{\text{ret}}]-\frac{\partial G_{\text{ret}}}{\partial\omega}\nabla_{\mathbf{k}}\Sigma^{<}\right]=\Sigma^{<}A-2\Gamma G^{<} \tag{8.290}$$

This equation can be simplified. The QBE is only valid to first power in the electric field, since terms of $O(\mathbf{E}^2)$ have systematically been ignored. For example, in the gradient expansion the second derivative terms would give contributions of $O(\mathbf{E}^2)$ were they retained. So the equation is exact to first order in the electric field, but is not exact to higher orders in the field. This fact is utilized to simplify the left-hand side of the equation. Since the field multiplies each term, on this side of the equal sign the Green's functions can be taken to have $\mathbf{E} = 0$. Of course, these expressions are just the equilibrium quantities in Sec. 3.7.

Consider the frequency derivatives of the left-hand side of the equation. Write the spectral function in the shorthand notation

$$A=\frac{2\Gamma}{\sigma^2+\Gamma^2},\qquad \Gamma=-\,\text{Im}[\Sigma_{\text{ret}}],\qquad \sigma=\omega-\varepsilon_{\mathbf{k}}-\text{Re}[\Sigma_{\text{ret}}] \tag{8.291}$$

The left-hand side will have three types of frequency derivatives: $\partial n_F/\partial\omega$, $\partial\Gamma/\partial\omega$, $\partial\sigma/\partial\omega$. The coefficients of the latter two vanish, which leaves only terms in $\partial n_F/\partial\omega$. All of the terms proportional to n_F vanish, leaving only terms multiplied by $\partial n_F/\partial\omega$:

$$A(\mathbf{k},\omega)^2\frac{\partial n_F}{\partial\omega}e\mathbf{E}\cdot\{(\mathbf{v}_{\mathbf{k}}+\nabla_{\mathbf{k}}\,\text{Re}[\Sigma_{\text{ret}}])\Gamma+\sigma\nabla_{\mathbf{k}}\Gamma\}=\Sigma^{>}G^{<}-\Sigma^{<}G^{>} \tag{8.292}$$

The factor of $A(\mathbf{k},\omega)^2$ appears in each term, and was taken outside. The left-hand side of this equation now contains only known quantities, which can be calculated in equilibrium. The scattering terms remain on the right-hand side. Finding them still involves work, usually in the form of an integral equation. This final form for the QBE is exact for transport which is linear in the field, and for steady state, homogeneous systems. It is quite analogous to the similar expression for the classical BE, which is

$$-e\mathbf{E}\cdot\mathbf{v}_{\mathbf{k}}\frac{\partial f^{(0)}}{\partial\omega}=\Sigma^{>}G^{<}-\Sigma^{<}G^{>} \tag{8.293}$$

The classical equation has $\partial f^{(0)}/\partial\omega$, while (8.292) has $\partial n_F/\partial\omega$.

Equation (8.292) is the steady state, homogeneous form of the QBE. It should be the starting point for many transport calculations. It is exact, and is an alternative to using the Kubo formalism, which is also exact. The derivation of this equation has been complicated, and has entailed some work. However, once derived, it is often the easiest starting point for deriving the transport coefficients. Calculations using Eq. (8.292) entail less work in getting to the answer than any other formalism.

8.5.3. Electron Scattering by Impurities

The quantum Boltzmann equation (QBE) (8.292) will be solved for the electron scattering by impurities. This case was solved in Sec. 8.1 using the Kubo formula. This exercise is useful for two reasons: (1) producing the known result demonstrates the correctness of the

QBE; and (2) this case is the easiest one to solve, and provides an introduction to the techniques for solving the QBE. The present example makes two assumptions that are intended to make the solution as easy as possible: (1) the impurities are dilute, so that the simultaneous scattering from several impurities can be neglected; and (2) the impurities have no internal degrees of freedom, such as spin or vibrations, which can be altered by the electron scattering. The second assumption implies that impurity is a simple potential that elastically scatters the electron. The impurities are randomly located in the solid. The method of Sec. 4.1.5 is employed for averaging over the random distributions.

The first step in solving the QBE is to find the retarded functions. For impurity scattering they are independent of the electric field, at least for small fields. The retarded self-energy for scattering from impurities of density n_i is given in Sec. 4.1

$$\Sigma_{\text{ret}}(\mathbf{k}, \omega) = n_i T_{\mathbf{kk}}(\omega) \tag{8.294}$$

The T matrix is energy dependent, and hence off-shell.

The next step is to express the self-energies $\Sigma^<$ and $\Sigma^>$ in terms of the Green's functions $G^<$ and $G^>$. The retarded functions G_{ret} and Σ_{ret} are known, since they are unchanged by the electric field. However, $G^<$ and $G^>$ are, at this point, unknown since they are affected by the electric field. The self-energy functions are

$$\Sigma^{<,>}(\mathbf{k}, \omega) = n_i \int \frac{d^3p}{(2\pi)^3} |T_{\mathbf{pk}}(\omega)|^2 G^{<,>}(\mathbf{p}, \omega) \tag{8.295}$$

This result is derived below. The off-diagonal T matrix is the one in Eq. (4.111), which also depends upon the energy ω.

Equation (8.295) is now derived. An impurity at $\mathbf{R} = 0$ is represented by an electron potential $V(\mathbf{r})$, whose Fourier transform is $V(\mathbf{q})$. The self-energy Σ from a single scattering event is

$$\tilde{\Sigma}_0(x_1, x_2) = V(\mathbf{r}_1)\delta^4(x_1 - x_2)\tilde{I} \tag{8.296}$$

where $\tilde{I}$ is the unit tensor. This self-energy is inserted into (2.157). That equation is iterated in order to find the effects of repeated scattering from the same impurity. The resulting self-energy series is rewritten in a symbolic notation, where the product of two functions implies an integral over dx. Iteration of these equations gives the series for Σ:

$$\tilde{\Sigma} = \tilde{\Sigma}_0[\tilde{I} + \tilde{G}_0\tilde{\Sigma}_0 + \tilde{G}_0\tilde{\Sigma}_0\tilde{G}_0\tilde{\Sigma}_0 + \cdots] \tag{8.297}$$

After summing this series, the matrix $\tilde{\Sigma}$ is examined for its individual components. The ones for $\Sigma^{<,>}$ are

$$\Sigma^{<,>} = [1 + G_{\text{ret}}\Sigma_{\text{ret}}]\Sigma_0^{<,>}[1 + \Sigma_{\text{adv}}G_{\text{adv}}] + \Sigma_{\text{ret}}G^{<,>}\Sigma_{\text{adv}} \tag{8.298}$$

Note the analogy with the equation for $G^<$ in (2.159). The resemblance is expected, since the series for Σ has the same mathematical structure as the one for G.

For impurity scattering, the unperturbed self-energies $\Sigma_0^{<,>} = 0$ are zero. There is only the last term in (8.298). Since the self-energies Σ_{ret} and Σ_{adv} are T matrices, (8.298) gives the result in (8.295). The damping function $\Gamma(\mathbf{k}, \omega)$ and other self-energy $\Sigma^>(\mathbf{k}, \omega)$ are

$$2\Gamma(\mathbf{k}, \omega) = n_i \int \frac{d^3p}{(2\pi)^3} |T_{\mathbf{pk}}|^2 A(\mathbf{p}, \omega) \tag{8.299}$$

$$\Sigma^> = \Sigma^< - 2i\Gamma \tag{8.300}$$

For dilute impurities, it is sufficient to retain only those terms that are first order in the impurity concentration n_i. The broadening due to the spectral function $A(\mathbf{p}, \omega)$ is from impurity scattering. Since the self-energy is multiplied by n_i, one can replace the spectral function $A(\mathbf{p}, \omega) \approx 2\pi\delta(\omega - \varepsilon_{\mathbf{p}})$. This expression is then equal to the imaginary part of the T matrix: $2\Gamma = n_i v_k \sigma_T$, where σ_T is the total cross section from impurity scattering in Sec. 4.1.

The starting point for solving the QBE is (8.292). On the left of the equals sign is the factor of

$$[(\mathbf{v}_{\mathbf{k}} + \nabla_{\mathbf{k}}\mathrm{Re}[\Sigma_{\mathrm{ret}}])\Gamma + \sigma\nabla_{\mathbf{k}}\Gamma]A(\mathbf{k}, \omega)^2 \tag{8.301}$$

This expression is simplified by neglecting terms of $O(n_i^2)$ such as $\Gamma\,\mathrm{Re}[\Sigma_{\mathrm{ret}}]$. The term in $\sigma = \omega - \varepsilon_{\mathbf{k}} - \mathrm{Re}[\Sigma_{\mathrm{ret}}]$ is small since the factor of $A(\mathbf{k}, \omega)^2$ tends to force $\sigma \approx 0$. The above expressions can be approximated by $v_{\mathbf{k}}\Gamma A(\mathbf{k}, \omega)^2$:

$$A(\mathbf{k}, \omega)^2 \frac{dn_F(\omega)}{d\omega} e\mathbf{E}\cdot\mathbf{v}_{\mathbf{k}}\Gamma = 2i\Gamma G^< - i\Sigma^< A \tag{8.302}$$

These steps complete the derivation of the left-hand side of the QBE.

On the right side of (8.302) the scattering terms vanish in equilibrium since $\Sigma^< = 2in_F\Gamma$ and $G^< = in_F A$. Since current is flowing in response to the field, the system is slightly out of equilibrium. The right-hand side is expected to have factors similar to those on the left. This discussion suggests the following ansatz for the nonequilibrium Green's function

$$G^< = iA(\mathbf{k}, \omega)\left[n_F(\omega) - \left(\frac{dn_F(\omega)}{d\omega}\right)e\mathbf{E}\cdot\mathbf{v}_{\mathbf{k}}\Lambda(\mathbf{k}, \omega)\right] \tag{8.303}$$

The function $\Lambda(\mathbf{k}, \omega)$ is unknown, and needs to be determined by solving the QBE. The factors that multiply Λ are for later convenience. The above choice does not make any assumptions for the value of Λ. Using this ansatsz in the self-energy function gives

$$\begin{aligned}\Sigma^< &= in_i \int \frac{d^3p}{(2\pi)^3}|T_{\mathbf{kp}}|^2 A(\mathbf{p}, \omega)\left[n_F(\omega) - \left(\frac{dn_F(\omega)}{d\omega}\right)e\mathbf{E}\cdot\mathbf{v}_{\mathbf{p}}\Lambda(\mathbf{p}, \omega)\right] \\ &= 2i\Gamma n_F - in_i\frac{dn_F}{d\omega}n_i\int\frac{d^3p}{(2\pi)^3}|T_{\mathbf{kp}}|^2 A(\mathbf{p}, \omega)e\mathbf{E}\cdot\mathbf{v}_{\mathbf{p}}\Lambda(\mathbf{p}, \omega)\end{aligned} \tag{8.304}$$

Putting these two expressions into the right-hand side of (8.302), the equilibrium terms cancel, and the remaining terms each have the common factor of $(dn_F/d\omega)A$:

$$\begin{aligned}A^2\frac{dn_F(\omega)}{d\omega}e\mathbf{E}\cdot v_{\mathbf{k}}\Gamma = -\left(\frac{dn_F(\omega)}{d\omega}\right)A\{&2\Gamma e\mathbf{E}\cdot\mathbf{v}_{\mathbf{k}}\Lambda \\ &- n_i\int\frac{d^3p}{(2\pi)^3}|T_{\mathbf{kp}}|^2 A(\mathbf{p}, \omega)e\mathbf{E}\cdot\mathbf{v}_{\mathbf{p}}\Lambda(\mathbf{p}, \omega)\}\end{aligned} \tag{8.305}$$

After canceling all of the common factors, there is an integral equation for the unknown function $\Lambda(\mathbf{k}, \omega)$:

$$\mathbf{v}_{\mathbf{k}}\Lambda(\mathbf{k}, \omega) = \tfrac{1}{2}A(\mathbf{k}, \omega)\mathbf{v}_{\mathbf{k}} + \frac{n_i}{2\Gamma}\int\frac{d^3p}{(2\pi)^3}|T_{\mathbf{kp}}|^2 A(\mathbf{p}, \omega)\mathbf{v}_{\mathbf{p}}\Lambda(\mathbf{p}, \omega) \tag{8.306}$$

This integral equation for $\Lambda(\mathbf{k}, \omega)$ is nearly identical to the one found in solving the conductivity from the Kubo relation. The quantity $\mathbf{v}_{\mathbf{k}}\Lambda(\mathbf{k}, \omega)$ is similar to the factor of

$\Gamma(\mathbf{k}, \omega - i\delta, \omega + i\delta)$ in Sec. 8.1.3. As the derivation proceeds, the differences between these quantities disappears, and the two approaches give the same resistivity.

After solving this equation for $\Lambda(\mathbf{k}, \omega)$, the Green's function $G^<$ in (8.303) is known. It is used in (3.340) for the current. The first term in (8.303) for $G^<$ gives a zero current. The second term gives a current proportional to the field $\mathbf{E}$, and this proportionality defines the electrical conductivity

$$\sigma_{\mu\nu} = e^2 \int \frac{d^3p}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} v_{p\mu} v_{p\nu} \left(-\frac{dn_F(\omega)}{d\omega} \right) A(\mathbf{p}, \omega)\Lambda(\mathbf{p}, \omega) \tag{8.307}$$

The two equations (8.306) and (8.307) provide the solution for the conductivity from the QBE.

For electrons in metals, with a spherical Fermi surface, the conductivity from impurity scattering is found easily from these equations. First, do the integral over $d\omega$ in (8.307). At low temperature the factor $(-dn_F/d\omega)$ is nearly a delta function which sets $\omega \approx 0$. At low temperature

$$\sigma_{\mu\nu} = e^2 \int \frac{d^3p}{(2\pi)^3} v_{\mathbf{p}\mu} v_{\mathbf{p}\nu} A(\mathbf{p}, 0)\Lambda(\mathbf{p}, 0) \tag{8.308}$$

The equation is reduced as far as possible. Next one must solve the equation (8.306) for Λ. Adopt a vector coordinate system where the $\hat{z}$ direction is $\hat{k}$. Then the various scalar products of vectors can be found from the law of cosines:

$$\hat{k} \cdot \hat{E} = \cos(\theta_0), \qquad \hat{k} \cdot \hat{p} = \cos(\theta) \tag{8.309}$$

$$\hat{p} \cdot \hat{E} = \cos(\theta_0)\cos(\theta) + \sin(\theta)\sin(\theta_0)\cos(\phi) \tag{8.310}$$

Equation (8.306) is multiplied by the vector $\hat{E}$. The integral over the $d\phi$ part of d^3p makes the $\cos(\phi)$ term vanish. Each of the remaining terms has a factor of $\cos(\theta_0)$ which can be canceled. Then one finds the scalar equation

$$\Lambda(k, 0) = \tfrac{1}{2}A(k, 0) + \frac{n_i}{2\Gamma} \int \frac{d^3p}{(2\pi)^3} |T_{\mathbf{kp}}|^2 A(\mathbf{p}, 0)\Lambda(\mathbf{p}, 0)\cos(\theta) \tag{8.311}$$

The spectral functions $A(k,0)$ and $A(p,0)$ force $k = p = k_F$. The factor of $\Lambda(p, 0)$ under the integral can be set equal to $\Lambda(k_F, 0)$ and taken out of the integral. Then the above equation can be solved to find

$$\Lambda(k_F, 0) = \frac{\tau_t}{2\tau} A(k_F, 0) \tag{8.312}$$

where $\tau = 1/2\Gamma$ is the time between scattering from impurities, while τ_t is the lifetime which is important for resistivity; it has the factor of $(1 - \cos\theta')$:

$$\sigma = \frac{ne^2\tau_t}{m} \tag{8.313}$$

$$\frac{1}{\tau_t} = n_i \int \frac{d^3p}{(2\pi)^3} |T_{\mathbf{kp}}|^2 A(\mathbf{p}, \omega)[1 - \cos(\theta)] \tag{8.314}$$

This formula for the resistivity is identical to (8.25). The QBE gives the same formula for the resistivity as found earlier from the Kubo formula. The two methods are also in exact agreement for much more complex cases, such as electron–phonon scattering at nonzero temperatures and frequency. Most of the steps in the derivation were spent in getting to

(8.302). This equation is the starting point for any calculation for homogeneous systems. The number of steps between this equation and the final resistivity is small. The QBE is an efficient method of finding the resistivity or other transport coefficients.

8.6. QUANTUM DOT TUNNELING

8.6.1. Electron Tunneling

Cohen *et al.* (1962) introduced the concept of the tunneling Hamiltonian, which has become universally adopted for the discussion of tunneling. Their idea was to write the Hamiltonian as three terms:

$$H = H_R + H_L + H_T \tag{8.315}$$

$$H_T = \sum_{\mathbf{kp}\sigma} (T_{\mathbf{kp}} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{p}\sigma} + \text{hc}) \tag{8.316}$$

The first term H_R is the Hamiltonian for particles on the right side of the tunneling junction. It contains all many-body interactions. Similarly, H_L has all the physics for particles on the left side of the junction. These two are considered to be strictly independent. Not only do these two operators commute, $[H_L, H_R] = 0$, but they commute term by term. The Hamiltonian on the right can be expressed in terms of one set of operators $C_{\mathbf{k}\sigma}$ and those on the left by another set $C_{\mathbf{p}\sigma}$, and these operators are independent $\{C_{\mathbf{k}\sigma}, C^\dagger_{\mathbf{p}\sigma}\} = 0$. This assumption is probably reasonable. They further assumed that the tunneling is caused by the term H_T in (8.316). The tunneling matrix element $T_{\mathbf{kp}}$ can transfer particles through an insulating junction. This transfer rate is assumed to depend only on the wave vectors on the two sides $\mathbf{k}$ and $\mathbf{p}$ and not on other variables, such as the energy or spin of the particles.

The theory of electron tunneling was mainly applied to superconductors. It developed very rapidly and was entirely based on the tunneling Hamiltonian. The theory showed excellent agreement with the many experiments. The history books were written describing this satisfactory situation, and the scientists in this field wandered off to do something else. About this time there began a serious investigation, starting with Zawadowski (1967), Caroli *et al.* (1975), and Feuchtwang (1975), about the validity of the tunneling Hamiltonian. Of course it was found to be a poor approximation, since the tunneling rate depends on the energy of the particle as well as its wave vector. What does this turn of events do to the lovely agreement between theory and experiment for tunneling in superconductors? Actually, it probably changes none of it. The tunneling in superconductors takes place over a very narrow span of energies in the metal, i.e., within a Debye energy of the Fermi surface. Also, all the electrons involved have their wave vector very near the Fermi wave vectors k_F and p_F on the two sides of the junction. It is an adequate approximation to treat the transfer rate $T_{\mathbf{kp}}$ as a constant T_0 which is evaluated at k_F and p_F, because the variations in $T_{\mathbf{kp}}$ with energy must be on the scale E/E_F which are negligible for $E \approx \Delta \approx 1$ meV. Similarly, the variation of $T_{\mathbf{kp}}$ with $\mathbf{p}$ or $\mathbf{k}$ is on the scale of the Fermi wave vectors. One can treat the transfer rate T_0 as a constant if the energies involved are small. The tunneling Hamiltonian is believed to be an improper formalism only when the applied voltages are large, say 1 eV.

For quantum dots the energy scales are also rather small, and are near to the chemical potentials of the metal electrodes. Here the tunneling Hamiltonian is also thought to be a valid formalism.

The general model of a tunneling junction is shown in Fig. 8.12. It describes a nonequilibrium situation, since the chemical potential on the left-hand side μ_L is not the same as μ_R on the right. They differ by the applied voltage $eV = \mu_L - \mu_R$. The potential drop of eV occurs in the insulating region between the electrodes.

The tunneling Hamiltonian (8.316) is used to derive a correlation function for electron tunneling currents. This correlation function has the form of a Kubo formula, except for an important difference. The Kubo formula for the conductivity in Sec. 3.7 expresses the ratio between the current and the voltage (actually electric field). In tunneling theory, the correlation function gives the current as a function of voltage.

The tunneling current through the insulating region is expressed as the rate of change of the number of particles on, for example, the left-hand side of the junction N_L. This rate is found from the commutator of $N_L = \sum C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma}$ with the tunneling Hamiltonian. Only the term H_T fails to commute with N_L,

$$\dot{N}_L = i[H, N_L] = i[H_T, N_L] = i \sum_{\mathbf{kp}\sigma} [T_{\mathbf{kp}} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{p}\sigma} - T^*_{\mathbf{kp}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{k}\sigma}] \tag{8.317}$$

The total current I through the tunneling interface is defined as the average value of this operator:

$$I(t) = -e\langle \dot{N}_L(t) \rangle \tag{8.318}$$

The average value of $\langle \dot{N}_L(t) \rangle$ is obtained by following the same steps used to derive the Kubo formula in Sec. 3.8. The total Hamiltonian is written as $H = H' + H_T$, where $H' = H_R + H_L$. Go to the interaction representation, where the tunneling term H_T is treated as the interaction and everything else H' is H_0. Then the S matrix is expanded in terms of the perturbation H_T. The objective is to obtain a formula where $I \propto |T_{\mathbf{kp}}|^2$, so only the first term needs to be retained in the expansion of the S matrix. These steps bring us to the formula

$$I(t) = -ei \int_{-\infty}^{t} dt' \langle [\dot{N}_L(t), H_T(t')] \rangle \tag{8.319}$$

$$H_T(t') = e^{iH't'} H_T e^{-iH't'} \tag{8.320}$$

$$\dot{N}_L(t) = e^{iH't} \dot{N}_L e^{-iH't} \tag{8.321}$$

where the time dependence of $H_T(t)$ and $\dot{N}_L(t)$ is governed by H'.

An important step in the calculation is to insert the chemical potentials μ_L and μ_R for the two sides of the junction. This insertion must be done with more care than usual, because the chemical potential is not the same on the two sides of the system. The initial Hamiltonian (8.316) has been written to not include the chemical potentials, so the energy is measured on

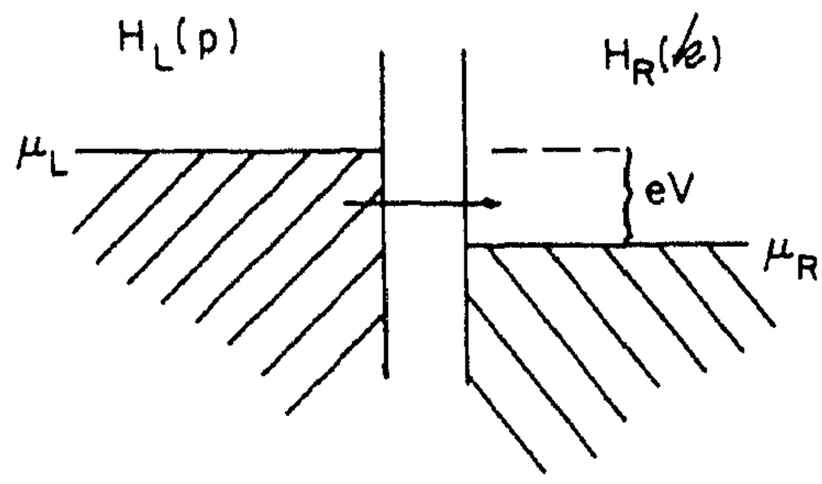


FIGURE 8.12 Tunneling between two normal metals. The arrow shows the electron path through the oxide interface

an absolute scale rather than relative to the chemical potentials. However, now insert the chemical potentials into the time developments, so that the energy can be measured with respect to the different chemical potentials on each side of the tunnel junction. The symbols K_R and K_L denote the Hamiltonian with respect to the respective chemical potentials:

$$K_R = H_R - \mu_R N_R \tag{8.322}$$

$$K_L = H_L - \mu_L N_L \tag{8.323}$$

$$K' = K_R + K_L \tag{8.324}$$

For a free-particle system, $K_R = \sum \xi_{\mathbf{k}} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{k}\sigma}$, while $H_R = \sum \varepsilon_{\mathbf{k}} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{k}\sigma}$ since $\xi_{\mathbf{k}} = \varepsilon_{\mathbf{k}} - \mu_R$. Since the number operators commute with H', it is possible to write $H' = K' + \mu_R N_R + \mu_L N_L$ and $\exp(iH't) = \exp(iK't)\exp[it(\mu_L N_L + \mu_R N_R)]$ since the exponentials can be separated when the operators commute. The time development of H_T is

$$\begin{aligned} H_T(t) &= e^{iH't} H_T e^{-iH't} = e^{iK't}\left[e^{it(\mu_L N_L + \mu_R N_R)} H_T e^{-it(\mu_R N_R + \mu_L N_L)}\right] e^{-itK'} \\ &= e^{iK't} \sum_{\mathbf{kp}\sigma} \left[T_{\mathbf{kp}} e^{it(\mu_R - \mu_L)} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{p}\sigma} + e^{it(\mu_L - \mu_R)} T^*_{\mathbf{kp}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{k}\sigma}\right] e^{-itK'} \end{aligned}$$

The commutator of H_T with the number operators produces the factor $\mu_L - \mu_R = eV$, which is identified as the applied voltage, as in Fig. 8.12. The applied voltage appears in the correlation function. The correlation function will now be evaluated by assuming that both sides of the junction are in separate thermodynamic equilibrium.

The current operator in (8.319) now becomes

$$\begin{aligned} I(t) = e\int_{-\infty}^{t} dt' \Big\langle &\sum_{\mathbf{kp}\sigma} [T_{\mathbf{kp}} e^{-ieVt} C^\dagger_{\mathbf{k}\sigma}(t) C_{\mathbf{p}\sigma}(t) - T^*_{\mathbf{kp}} e^{ieVt} C^\dagger_{\mathbf{p}\sigma}(t) C_{\mathbf{k}\sigma}(t)], \\ &\sum_{\mathbf{k}'\mathbf{p}'\sigma'} [T_{\mathbf{k}'\mathbf{p}'} e^{-ieVt'} C^\dagger_{\mathbf{k}'\sigma'}(t') C_{\mathbf{p}'\sigma'}(t') + T^*_{\mathbf{k}'\mathbf{p}'} e^{ieVt'} C^\dagger_{\mathbf{p}'\sigma'}(t') C_{\mathbf{k}'\sigma'}(t')]\Big\rangle \end{aligned}$$

From now on the time development of $C_{\mathbf{k}\sigma}$ operators is governed by $C_{\mathbf{k}\sigma}(t) = e^{iK_R t} C_{\mathbf{k}\sigma} e^{-iK_R t}$ and $C_{\mathbf{p}\sigma}$ operators by $C_{\mathbf{p}\sigma}(t) = e^{iK_L t} C_{\mathbf{p}\sigma} e^{-iK_L t}$. Define the operator A as

$$A(t) = \sum_{\mathbf{kp}\sigma} T_{\mathbf{kp}} C^\dagger_{\mathbf{k}\sigma}(t) C_{\mathbf{p}\sigma}(t) \tag{8.325}$$

and the current is written as the summation of two currents

$$I = I_S + I_J \tag{8.326}$$

$$I_S(t) = e\int_{-\infty}^{\infty} dt' \Theta(t - t')\{e^{ieV(t'-t)}\langle[A(t), A^\dagger(t')]\rangle - e^{ieV(t-t')}\langle[A^\dagger(t), A(t')]\rangle\} \tag{8.327}$$

$$I_J(t) = e\int_{-\infty}^{\infty} dt' \Theta(t - t')\{e^{-ieV(t'+t)}\langle[A(t), A(t')]\rangle - e^{ieV(t+t')}\langle[A^\dagger(t), A^\dagger(t')]\rangle\} \tag{8.328}$$

The term I_S is for single-particle tunneling, which is important for quantum dots. The other term I_J describes the tunneling currents associated with the Josephson effect in superconductors. It is evaluated in Chapter 10.

The terms in I_S have just the right combination of factors to be a retarded Green's function. The integrand depends only on the difference of$(t - t')$, so set $t' = 0$:

$$\bar{U}_{\text{ret}}(t) = -i\Theta(t)\langle[A(t), A^{\dagger}(0)]\rangle \tag{8.329}$$

$$U_{\text{ret}}(-eV) = \int_{-\infty}^{\infty} dt e^{-ieVt}\bar{U}_{\text{ret}}(t) \tag{8.330}$$

The second term in I_S is the Hermitian conjugate of $U_{\text{ret}}(-eV)$ except for a sign. But the Hermitian conjugate of the retarded function is just the advanced function. The single-particle tunneling current is written as;

$$I_S = ie[U_{\text{ret}}(-eV) - U_{\text{adv}}(-eV)] \tag{8.331}$$

$$= -2e\ \text{Im}[U_{\text{ret}}(-eV)] \tag{8.332}$$

It is twice the imaginary part of a retarded correlation function, which has the form of a spectral density function. Using the relationship between the Matsubara and the retarded correlation functions discussed in Sec. 3.3, the way to calculate the single-particle tunneling is to evaluate in the Matsubara formalism the correlation function

$$\begin{aligned} U(i\omega) &= -\int_0^{\beta} d\tau e^{i\omega_n\tau}\langle T_\tau A(\tau)A^{\dagger}(0)\rangle \\ &= -\sum_{\mathbf{kp}\sigma}\sum_{\mathbf{k'p'}\sigma'} T_{\mathbf{kp}}T^*_{\mathbf{k'p'}}\int_0^{\beta} d\tau e^{i\omega_n\tau}\langle T_\tau C^{\dagger}_{\mathbf{k}\sigma}(\tau)C_{\mathbf{p}\sigma}(\tau)C^{\dagger}_{\mathbf{p'}\sigma'}C_{\mathbf{k'}\sigma'}\rangle \end{aligned} \tag{8.333}$$

The single-particle tunneling current is just the spectral function of this operator evaluated at the real frequency $-eV/\hbar$, as shown in (8.330). The Matsubara frequency $\omega_n = 2n\pi/\beta$ is boson, since the correlation function has pairs of fermion operators. In the tunneling Hamiltonian, the right- and left-hand sides of the tunneling junction are independent. Then the correlation function factors into a product of the Green's functions for the right and left sides of the junction:

$$\begin{aligned} U(i\omega) &= \sum_{\mathbf{kp}\sigma}|T_{\mathbf{kp}}|^2\int_0^{\beta} d\tau e^{i\omega_n\tau}\langle T_\tau C_{\mathbf{k}\sigma}(0)C^{\dagger}_{\mathbf{k}\sigma}(\tau)\rangle\langle T_\tau C_{\mathbf{p}\sigma}(\tau)C^{\dagger}_{\mathbf{p}\sigma}(0)\rangle \\ &= \sum_{\mathbf{kp}\sigma}|T_{\mathbf{kp}}|^2\int_0^{\beta} d\tau e^{i\omega_n\tau}\mathcal{G}_R(\mathbf{k}, -\tau)\mathcal{G}_L(\mathbf{p}, \tau) \\ &= \sum_{\mathbf{kp}\sigma}|T_{\mathbf{kp}}|^2\frac{1}{\beta}\sum_{ip}\mathcal{G}_R(\mathbf{k}, ip - i\omega)\mathcal{G}_L(\mathbf{p}, ip) \end{aligned} \tag{8.334}$$

The only terms which enter are those for $\mathbf{k} = \mathbf{k}'$, $\mathbf{p} = \mathbf{p}$ and $\sigma = \sigma'$. A type of Feynman diagram is shown in Fig. 8.13. The symbols T in circles are the tunneling vertices. The vertical dashed line is meant to divide the right from the left side of the junction. The solid lines are the interacting Green's functions $\mathcal{G}_R$ and $\mathcal{G}_L$.

The correlation function is the product of the two Green's functions for the two sides of the junction. Such a simple result is obtained only when one can neglect the vertex diagrams of the correlation function. The argument for neglecting them in tunneling is that vertex corrections require some interaction between electrons on the two sides of the junction, which is improbable. Actually, there is another case where vertex corrections are needed in

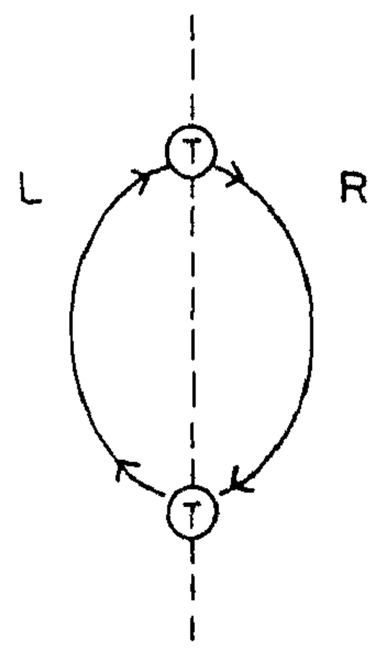


FIGURE 8.13 Feynman diagram for tunneling, where the Ts are the vertices which link the right (R) and left (L) sides of the junction.

tunneling, and that is to account for processes where the tunneling electron can excite vibrations or other excitations in the interface. That happens in quantum dot tunneling.

The first example of solving (8.334) will be for the tunneling between two normal metals. The experiments show the tunneling is linear in voltage V for small values of V, and becomes slowly nonlinear at larger values of V. The latter behavior is a failure of the tunneling Hamiltonian, which predicts only linear behavior. The present discussion is limited to small voltages, for which the tunneling Hamiltonian formalism is valid.

For a normal system, consider the electrons to be simple quasiparticles, and approximate the Green's functions by $\mathscr{G}_R^{(0)}$ and $\mathscr{G}_L^{(0)}$. The Matsubara summation is then familiar:

$$\frac{1}{\beta}\sum_{ip}\mathscr{G}_L^{(0)}(\mathbf{p}, ip)\mathscr{G}_R^{(0)}(\mathbf{k}, ip - i\omega) = \frac{n_F(\xi_k) - n_F(\xi_\mathbf{p})}{i\omega + \xi_\mathbf{k} - \xi_\mathbf{p}} \tag{8.335}$$

$$U(i\omega) = \sum_{\mathbf{kp}\sigma}|T_\mathbf{kp}|^2\frac{n_F(\xi_\mathbf{k}) - n_F(\xi_\mathbf{p})}{i\omega + \xi_\mathbf{k} - \xi_\mathbf{p}} \tag{8.336}$$

$$\begin{aligned} I &= -2e\,\mathrm{Im}[U_{\mathrm{ret}}(eV)] \\ &= 4\pi e\sum_{\mathbf{kp}}|T_\mathbf{kp}|^2[n_F(\xi_\mathbf{k}) - n_F(\xi_\mathbf{p})]\delta(eV + \xi_\mathbf{k} - \xi_\mathbf{p}) \end{aligned}$$

The energies $\xi_\mathbf{k}$ and occupation factor $n_F(\xi_\mathbf{k})$ refer to the right side of the junction, and $\xi_\mathbf{p}$ and $n_F(\xi_\mathbf{p})$ refer to the left side. An additional factor of two in front is for the summation over spins. The spin is usually preserved in tunneling. The summations over $\mathbf{k}$ and $\mathbf{p}$ are just over wave vectors. For small voltages, only electrons very near to the Fermi energy are involved in tunneling. For these small energies, the density of states on both sides is assumed to be a constant:

$$\sum_\mathbf{k} \to \int\frac{d^3k}{(2\pi)^3} \to N_R\int d\xi_R \tag{8.337}$$

$$\sum_\mathbf{p} \to \int\frac{d^3p}{(2\pi)^3} \to N_L\int d\xi_L \tag{8.338}$$

$$\begin{aligned} I &= 4\pi e N_R N_L|T|^2\int d\xi_L\int d\xi_R[n_F(\xi_R) - n_F(\xi_L)]\delta(eV + \xi_R - \xi_L) \\ &= 4\pi e N_R N_L|T|^2\int d\xi_R[n_F(\xi_R) - n_F(\xi_R + eV)] \end{aligned} \tag{8.339}$$

At zero temperature, the occupation numbers are step functions, so (8.339) becomes

$$I = 4\pi e N_R N_L |T|^2 \int_{-eV}^{0} d\xi_R = \sigma_0 V \tag{8.340}$$

$$\sigma_0 = 4\pi e^2 N_R N_L |T|^2 \tag{8.341}$$

The tunneling device behaves as a simple resistor, with a conductance (inverse resistance) given by σ_0. With a little more care, one can show that the integral in (8.339) also equals eV at nonzero temperatures, since the thermal smearing cancels between $n_F(\xi)$ and $n_F(\xi + eV)$. The result is temperature independent, at least for a range of low temperatures $k_B T \ll E_F$.

A formal expression for the tunneling current can also be derived when interacting Green's functions $\mathcal{G}_R$, $\mathcal{G}_L$ are retained in the Matsubara summation. This type of summation was evaluated in Sec. 7.1 using the Lehmann representation

$$\begin{aligned} X(i\omega) &= \frac{1}{\beta}\sum_{ip} \mathcal{G}_L(\mathbf{p}, ip)\mathcal{G}_R(\mathbf{k}, ip - i\omega) \\ &= \int \frac{d\varepsilon_L}{2\pi} A_L(\mathbf{p}, \varepsilon_L) \int \frac{d\varepsilon_R}{2\pi} A_r(\mathbf{k}, \varepsilon_R) S \end{aligned} \tag{8.342}$$

$$S = \frac{1}{\beta}\sum_{ip} \frac{1}{ip - \varepsilon_L} \frac{1}{ip - i\omega - \varepsilon_R} = \frac{n_F(\varepsilon_R) - n_F(\varepsilon_L)}{i\omega + \varepsilon_R - \varepsilon_L} \tag{8.343}$$

The retarded function is obtained by the analytical continuation $i\omega \to eV + i\delta$ and then take the imaginary part.

$$-2\,\mathrm{Im}[X_{\mathrm{ret}}(eV)] = \int \frac{d\varepsilon_R}{2\pi} A_L(\mathbf{p}, \varepsilon_R + eV) A_R(\mathbf{k}, \varepsilon_R)[n_F(\varepsilon_R) - n_F(\varepsilon_R + eV)]$$

These steps bring us to the formula for the tunneling current of Schrieffer *et al.* (1963):

$$I = 2e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \int \frac{d\varepsilon_R}{2\pi} A_L(\mathbf{p}, \varepsilon_R + eV) A_R(\mathbf{k}, \varepsilon_R)[n_F(\varepsilon_R) - n_F(\varepsilon_R + eV)] \tag{8.344}$$

Again a factor of two is added for the spin summation. The tunneling current is expressed in terms of the spectral functions on the two sides of the junction. Equation (8.344) is the exact formula for I within the model of the tunneling Hamiltonian. The earlier expression (8.339) is recovered with the free-quasiparticle approximation $A_R = 2\pi\delta(\varepsilon_R - \xi_{\mathbf{k}})$ and $A_L = 2\pi\delta(\varepsilon_L - \xi_{\mathbf{p}})$. The virtue of (8.344) is that it contains all many-body effects on the two sides of the junction. Its drawback is that it is based on the tunneling Hamiltonian formalism.

If the tunneling matrix element $T_{\mathbf{kp}}$ can be approximated as a constant, then the summations over wave vectors give the interacting density of states

$$\sum_{\mathbf{k}} A_R(\mathbf{k}, \varepsilon_R) = 2\pi N_R(\varepsilon_R) \tag{8.345}$$

$$\sum_{\mathbf{p}} A_L(\mathbf{p}, \varepsilon_L) = 2\pi N_L(\varepsilon_L) \tag{8.346}$$

$$I = 4\pi e |T|^2 \int d\varepsilon N_R(\varepsilon) N_L(\varepsilon + eV)[n_F(\varepsilon) - n_F(\varepsilon + eV)]$$

which is a useful formula for describing many-body effects in tunneling which occur on a small energy scale near to the Fermi surface.

A useful formula for quantum dots is to calculate the rate of electron tunneling in each direction: left-to-right ("LR") and right-to-left ("RL"):

$$I_{RL} = 2e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \int \frac{d\varepsilon_R}{2\pi} A_L(\mathbf{p}, \varepsilon_R + eV) A_R(\mathbf{k}, \varepsilon_R) n_F(\varepsilon_R)[1 - n_F(\varepsilon_R + eV)]$$

$$\approx 4\pi e |T|^2 N_R N_L \int d\varepsilon n_F(\varepsilon)[1 - n_F(\varepsilon + eV)] = \frac{\sigma_0 V}{1 - e^{-\beta eV}} \tag{8.347}$$

$$I_{LR} = 2e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \int \frac{d\varepsilon_R}{2\pi} A_L(\mathbf{p}, \varepsilon_R + eV) A_R(\mathbf{k}, \varepsilon_R) n_F(\varepsilon_R + eV)[1 - n_F(\varepsilon_R)]$$

$$\approx 4\pi e |T|^2 N_R N_L \int d\varepsilon n_F(\varepsilon + eV)[1 - n_F(\varepsilon)] = \frac{\sigma_0 V}{e^{\beta eV} - 1} \tag{8.348}$$

$$I_{RL} - I_{LR} = \sigma_0 V \tag{8.349}$$

The previous result is obtained by subtracting the currents in the two directions.

8.6.2. Quantum Dots

Figure 8.14 shows the typical geometry with a quantum dot (QD). There are two metal electrodes on the right and left, and a spherical metal particle between them. The sphere is imbedded in an insulating material through which the electron can tunnel. The tunneling of electrons from the right to the left electrode is facilitated by the QD. The tunneling becomes a two-step process: the electron tunnels from one electrode to the QD, and then tunnels from the QD to the other electrode. There is also coherent tunneling from one electrode to the other through the states of the QD, which act as virtual intermediate states. The first experiments of this type were done by Giaever and Zeller (1968), and they have become popular recently. Now the QD and electrodes might be composed of semiconductors, where the nanostructure is made by lithography.

First examine the properties of the QD. Assume it is a sphere of radius a in atomic units: the dimensional radius $R = aa_0$ where $a_0 = 0.05292\,\text{nm}$ is the Bohr radius. The volume of the sphere is $V_0 = 4\pi R^3/3$. The number of electrons is $N = n_0 V_0$, where the density

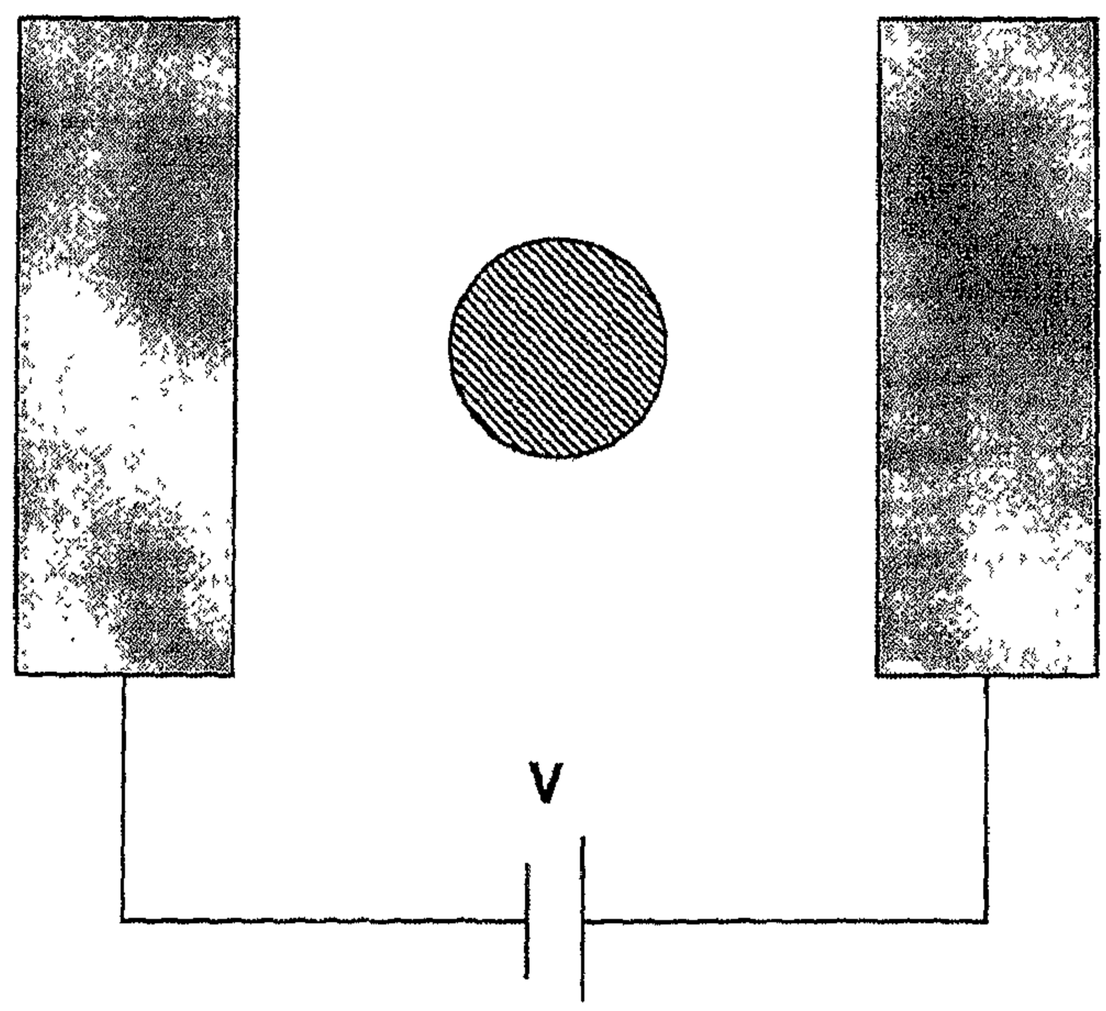


FIGURE 8.14 Geometry for quantum dot tunneling.

$n_0 = k_F^3/3\pi^2$ is assumed to be the same as in the bulk metal. The density of states per spin is $N_F = mk_F/(2\pi^2\hbar^2)$. The quantity $N_F V_0$ has the units of inverse energy: it is the number of electrons per unit energy. Conversely, its inverse is the average separation between energy levels on the QD

$$\Delta E = \frac{1}{N_F V_0} = 3\pi \frac{E_{ry}}{(k_F a_0)a^3} \tag{8.350}$$

$$U = eV_c = \frac{e^2}{2R} = \frac{E_{ry}}{a} \tag{8.351}$$

The quantity U is called the *charging energy*. It is the energy required to add a charge of e to a neutral metal sphere. All of the charge collects on the surface, and the energy is just $U = e^2/2R$. Table 8.3 compares these two energies as a function of sphere radius. The energies are given in terms of electron volts. The two energy scales are interesting. The charging energy is large compared to the thermal energy $k_B T$, regardless of which temperature is used. If $\bar{n}$ is the number of electrons on the QD for it to be charge neutral, then the energy for n electrons on the QD is

$$E_n = \Delta E(n - \bar{n}) + U(n - \bar{n})^2 \tag{8.352}$$

The system behaves as an Anderson model, in that the energy of the QD depends quadratically upon the total number of electrons n. The energy is for a many-electron system. The energy U is due to electron–electron interactions. A many-electron description is required for treating the QD.

The other energy ΔE is also interesting. It is much smaller than the charging energy for typical systems with $N \sim 10$–1000 electrons. Here the relevent question is whether ΔE is greater or smaller than the thermal energy $k_B T$. Of course, this comparison depends upon the temperature of the experiment: whether it is at 300 K, 1 K, or 1 mK. If $\beta\Delta E \ll 1$ then the energy levels on the QD can be treated as a continuum, and no error is made by replacing the summation over states by a continuous integral. However, if $\beta\Delta E \gg 1$, the states are well separated in energy compared to $k_B T$. The summation over states must remain a discrete summation. Either condition could apply, depending upon the size of the QD and the temperature of the experiment. Different theories in the literature assume one case or the other. Some simple statistical calculations are presented of the fluctuation of the number of

Table 8.3 Number of electrons N, charging energy U, and energy level separation ΔE of a quantum dot of radius a in atomic units. Energies in units of electron volts. A value of $r_s = 3.96$ (sodium) is assumed

a	N	U	ΔE
5	2	2.72	2.116
10	16	1.36	0.265
15	54	0.91	0.078
20	129	0.68	0.033
25	252	0.54	0.017
30	435	0.45	0.010
35	690	0.39	0.006
40	1030	0.34	0.004

electrons on quantum dots. These fluctuations are important for controlling, or else not controlling, the flow of current onto the quantum dot. These fluctuations play a role in the phenomena called "Coulomb Blockade". The two different formulas in the literature for the tunneling rate are obtain as two limits of the fluctuations.

The statistical averaging over the states n of the quantum dot includes both the single-electron states $(n-\bar{n})\Delta E$ and the charging energy $U(n-\bar{n})^2$. Sometimes they can be decoupled. There are, in fact, many configurations of one-electron energy states. Denote their quantum numbers by α, so that the energy without the charging energy is E_α.

Calculate the partition function for the electrons on the quantum dot. Assume there are a number of single-electron energy levels with energy E_α as well as a charging energy $U(n-\bar{n})^2$. Let n_α be zero or one, and it denotes whether an electron is, or is not, in the state α. Then the total energy of the system is

$$n = \sum_\alpha n_\alpha \tag{8.353}$$

$$E = \sum_\alpha n_\alpha E_\alpha + U(n-\bar{n})^2 \tag{8.354}$$

The label α includes the spin and other quantum numbers. Many combinations of states may have the same value of n. In this case, the partition function can be written exactly as

$$Z = \sum_n e^{-\beta U(n-\bar{n})^2} \int_0^{2\pi} \frac{d\theta}{2\pi} e^{in\theta} \Pi_\alpha \sum_{n_\alpha=0}^{1} e^{-n_\alpha(\beta(E_\alpha-\mu)+i\theta)} \tag{8.355}$$

$$= \sum_n e^{-\beta U(n-\bar{n})^2} \int_0^{2\pi} \frac{d\theta}{2\pi} e^{in\theta} S(\theta) \tag{8.356}$$

$$S(\theta) = \Pi_\alpha[1 + e^{-\beta(E_\alpha-\mu)-i\theta}] \tag{8.357}$$

For $\theta = 0$ then $S(0)$ is just the usual Fermi–Dirac distribution function. The integral over θ selects out the terms with a certain value of n. If $U = 0$ then the summation over n gives a delta function at $\theta = 0$ and the partition function reverts to being a usual product of Fermi–Dirac distributions for noninteracting electrons.

The role of the fluctuations is determined by the function $S(\theta)$. This function is evaluated by breaking the summation over α into two parts depending upon whether $\xi_\alpha = E_\alpha - \mu$ is positive or negative, i.e., whether $\exp(-\beta\xi_\alpha)$ is less than or larger than one.

$$\ln(S(\theta)) = \sum_{\xi_\alpha<0} \ln[e^{-i\theta}(e^{i\theta} + e^{-\beta\xi_\alpha})] + \sum_{\xi_\alpha>0} \ln\left[(1+e^{-\beta\xi_\alpha})\left(1 + \frac{e^{-\beta\xi_\alpha}(e^{-i\theta}-1)}{1+e^{-\beta\xi_\alpha}}\right)\right]$$

$$= \ln[S(0)] - i\theta\bar{n} + \delta S(\theta) \tag{8.358}$$

$$\bar{n} = \sum_{\xi_\alpha<0} 1 \tag{8.359}$$

$$\tilde{n}_\alpha = \frac{1}{e^{\beta\xi_\alpha}+1} = n_F(\xi_\alpha) \tag{8.360}$$

$$\delta S(\theta) = \sum_{\xi_\alpha<0} \ln[1 + (1-\tilde{n}_\alpha)(e^{i\theta}-1)] + \sum_{\xi_\alpha>0} \ln[1 + \tilde{n}_\alpha(e^{-i\theta}-1)] \tag{8.361}$$

where $\bar{n}$ is the number of electrons below the chemical potential, and $\tilde{n}_\alpha$ is the Fermi–Dirac distribution of electrons for the state α.

First evaluate δS for the case that the energy separation between the states of energy E_α is less than the thermal energy $k_B T = 1/\beta$. Then the summation over α can be replaced by a continuous integral. Let $N_F(\xi)V_0 = \rho_D(\xi)$, $\rho_D(0) = 1/\Delta E$ be the density of energy states on the dot. Then one gets that

$$\sum_{\xi_\alpha} = V_0 \int N_F(\xi) d\xi = \int d\xi \rho_D(\xi) \tag{8.362}$$

$$\delta S = \int_{-W}^{0} d\xi \rho_D(\xi) \ln[1 + n_F(-\xi)(e^{i\theta} - 1)] + \int_0^W d\xi \rho_D(\xi) \ln[1 + n_F(\xi)(e^{-i\theta} - 1)] \tag{8.363}$$

$$\delta S = -\sum_{l=1} \alpha_l [1 - \cos(l\theta)] \tag{8.364}$$

$$\delta S \approx -2\gamma[1 - \cos(\theta)] \tag{8.365}$$

$$\gamma = k_B T \rho_D(0) \ln 2 = \frac{k_B T}{\Delta E} \ln(2) \tag{8.366}$$

The integral has been approximated by expanding the log function in a Taylor series and keeping the first term. Higher terms ($l > 1$) in the series for δS are much smaller: the series converges rapidly. In general $\gamma > 1$ means that $k_B T > \Delta E$.

With this approximation the partition function can be expressed in closed form

$$Z = Z_{FD} Z_C \tag{8.367}$$

$$Z_{FD} = \Pi_\alpha [1 + e^{-\beta \xi_\alpha}] \tag{8.368}$$

$$Z_C = \sum_\nu \mathscr{I}_\nu(\gamma) e^{-\beta U \nu^2} \tag{8.369}$$

$$\nu = n - \bar{n} \tag{8.370}$$

$$\mathscr{I}_\nu(\gamma) = \int_0^{2\pi} \frac{d\theta}{2\pi} e^{i\nu\theta - 2\gamma(1-\cos(\theta))} \tag{8.371}$$

$$= e^{-2\gamma} I_\nu(2\gamma) \approx \frac{1}{2\sqrt{\pi\gamma}} e^{-\nu^2/4\gamma} \tag{8.372}$$

This formula is an important result. It shows that the Fermi–Dirac statistics of the states labeled α are decoupled from the charging of the QD, which is given by the last term Z_C. When $\gamma > 1$, then for small values of ν the factor of $\mathscr{I}_\nu(\gamma)$ is independent of ν, and the averaging over the value of n involves only the term $\exp(-\beta U \nu^2)$. This result is assumed by most theoretical papers. The charging is treated as a separate statistical process from the Fermi–Dirac statistics of the electrons on the QD.

The other case is to evaluate $\delta S(\theta)$ while assuming that the separation between energy levels is greater than the thermal energy. In that case this function is evaluated as

$$\delta S \approx -\gamma^{<}(1 - e^{i\theta}) - \gamma^{>}(1 - e^{-i\theta}) \tag{8.373}$$

$$\gamma^{<} = \sum_{\xi_\alpha < 0} n_F(-\xi_\alpha) \tag{8.374}$$

$$\gamma^{>} = \sum_{\xi_\alpha > 0} n_F(\xi_\alpha) \tag{8.375}$$

$$\mathscr{I}_\nu = e^{-(\gamma^{<}+\gamma^{>})}\left(\frac{\gamma^{>}}{\gamma^{<}}\right)^{\nu/2} I_\nu(2\sqrt{\gamma^{<}\gamma^{>}}) \tag{8.376}$$

$$Z = Z_{FD}Z_C \tag{8.377}$$

$$Z_C = \sum_\nu \mathscr{I}_\nu e^{-\beta U \nu^2} \tag{8.378}$$

Here $\gamma < 1$. The formulas for the two cases become identical when $\gamma = \gamma^{<} = \gamma^{>}$.

However, if $\gamma < 1$ then $\mathscr{I}_\nu$ does depend upon γ and this factor should be included in the analysis. Our derivation shall assume that $\gamma > 1$, so that the energy levels on the QD can be considered to be continuous. The thermodynamic averaging over the charging energy is accomplished by including the factor of $\exp(-\beta U n^2)$. In this case, the system is a type of Hubbard model.

8.6.3. Rate Equations

The Coulomb blockade is derived using a set of rate equation, which describe how an electron hops on or off the QD. Define f_n as the probability of there being n electrons on the QD. If the number is $\bar{n}$ in equilibrium, then we expect the probability to be something like

$$f_n = \frac{1}{Z_{QD}} \exp[-\beta(n - \bar{n})\Delta - \beta(n - \bar{n})^2 U] \tag{8.379}$$

$$1 = \sum_n f_n \tag{8.380}$$

where Δ is the energy to put a single electron on the QD, irrespective of the charging energy. It is convenient to change the notation and to define $n' = n - \bar{n}$ as the number of excess electrons ($n' > 0$) or holes ($n' < 0$) on the QD. Similarly, the label on the site probability is changed to $f_{n'}$. The subscript denotes the deviations from neutrality.

Define $R_L(m, n)$ as the probability that the number of electrons on the QD changes from n to m by a jump to the electrode to the left. Similarly, $R_R(m, n)$ is the probability of going

from n to m by a jump to the electrode on the right. Since only single jumps are considered, then $m = n \pm 1$. These quantities are given earlier as $I_{RL,LR}$ in Eq. (8.347) and (8.348).

$$R_L(n+1,n) = \frac{\xi_L E_L(n)}{e^{\beta E_L(n)} - 1} \tag{8.381}$$

$$R_L(n-1,n) = \frac{\xi_L E_L(n-1)}{1 - e^{-\beta E_L(n-1)}} \tag{8.382}$$

$$R_R(n+1,n) = \frac{\xi_R E_R(n)}{e^{\beta E_R(n)} - 1} \tag{8.383}$$

$$R_R(n-1,n) = \frac{\xi_R E_R(n-1)}{1 - e^{-\beta E_R(n-1)}} \tag{8.384}$$

$$\xi_{R,L} = 4\pi |T_{L,R}|^2 N_{L,R} N_{QD} \tag{8.385}$$

$$E_L(n) = eV_{QD} + U(2n+1) \tag{8.386}$$

$$E_R(n) = e(V_{QD} - V) + U(2n+1) \tag{8.387}$$

The left electrode is taken to have zero voltage, while the right one has a voltage of V. The quantum dot has a voltage of V_{QD}. The voltage differences V_{QD} and $V_{QD} - V$ are the differences in the chemical potential between the left electrode and the QD, and between the QD and the right electrode. The factor of $(2n+1) = (n+1)^2 - n^2$ is the difference between the charging energies before and after a hop that adds an electron to the QD. The factors of N_L, N_R, N_{QD} are the density of states at the chemical potential for the left and right electrodes, and for the QD.

As an example, consider the subsystem of just the QD and the left electrode. The rate equation for the process of having electrons hop between the QD and the electrode is

$$\frac{d}{dt} f_n = -f_n[R_L(n+1,n) + R_L(n-1,n)] + f_{n+1} R_L(n,n+1) + f_{n-1} R_L(n,n-1) \tag{8.388}$$

The first term on the right of the equals sign is from events where the system starts in state n and changes by having an electron hop either on to or off of the QD. The remaining two terms are events where the QD starts with either $n \pm 1$ electrons and gets to n by a hop. In equilibrium, the average rate of change is zero. Set the right-hand side of the above equation to zero. This step can be accomplished with the identities

$$f_n R_L(n+1,n) = f_{n+1} R_L(n,n+1) \tag{8.389}$$

$$f_{n-1} R_L(n,n-1) = f_n R_L(n-1.n) \tag{8.390}$$

Adding these two equation gives the right-hand side of (8.388). However, the above two equations are actually identical. The second is the same as the first by changing $n \to n-1$ everywhere. Only consider the solution to the first equation, which can be manipulated to write it as

$$\frac{f_{n+1}}{f_n} = \frac{R_L(n+1,n)}{R_L(n,n+1)} = \exp[-\beta E_L(n)] \tag{8.391}$$

$$f_n = f_0 \exp[-eV_{QD} n - \beta U n^2] \tag{8.392}$$

The second equation above is the solution deduced from the first. The result is expected for thermal equilibrium.

The case for the QD tunneling is different. There is a voltage V across the device, and a current flowing, so that the system is not in thermal equilibrium. The relationship among the factors of f_n must be derived anew. The rate equation for this case, including both electrodes, is

$$\begin{aligned}\frac{df_n}{dt} = &-f_n[R_L(n+1,n) + R_L(n-1,n) + R_R(n+1,n)\\ &+ R_R(n-1,n)] + f_{n+1}[R_L(n,n+1) + R_R(n,n+1)]\\ &+ f_{n-1}[R_L(n,n-1) + R_R(n,n-1)]\end{aligned} \tag{8.393}$$

In steady state, the average derivative is zero, and the right-hand side vanishes. As in the above case, this constraint is accomplished by

$$\begin{aligned}f_n[R_L(n+1,n) + R_R(n+1,n)] &= f_{n+1}[R_L(n,n+1) + R_R(n,n+1)]\\ \frac{f_{n+1}}{f_n} &= \frac{R_L(n+1,n) + R_R(n+1,n)}{R_L(n,n+1)R_R(n,n+1)}\end{aligned} \tag{8.394}$$

The above equation determines the occupation numbers f_n for the quantum dot. Another important equation is for the current. One can count either the electrons leaving the first electrode, or else those arriving at the second one. These two expressions for the current are

$$\begin{aligned}I_L &= e\sum_n f_n[R_L(n+1,n) - R_L(n-1,n)]\\ &= e\sum [f_n R_L(n+1,n) - f_{n+1}R_L(n,n+1)]\end{aligned} \tag{8.395}$$

$$\begin{aligned}I_R &= e\sum_n f_n[R_R(n-1,n) - R_R(n+1,n)]\\ &= e\sum_n [f_{n+1}R_R(n,n+1) - f_n R_R(n+1,n)]\end{aligned} \tag{8.396}$$

In the second equal sign we changed $n \to n+1$ in one of the terms, which is permitted since there is a summation over n. In steady state the two currents are identical: $I_L = I_R$. This identity is automatically satisfied by Eq. (8.394).

As a simple example, consider the case that the charging energy $U = 0$. In that case the solution to the rate equations is that

$$\xi_L E_L = -\xi_R E_R \tag{8.397}$$

$$\xi_L e V_{QD} = -\xi_R e(V_{QD} - V) \tag{8.398}$$

$$V_{QD} = V\frac{\xi_R}{\xi_R + \xi_L} \tag{8.399}$$

Putting this result into Eq. (8.394) gives that the ratio $f_{n+1}/f_n = 1$. All of the occupation numbers f_n are the same. An evaluation of the current gives that

$$I = -e^2 V\frac{\xi_L \xi_R}{\xi_L + \xi_R} \tag{8.400}$$

This result is the same as that obtained by two resistors in series. A simple tunnel junction acts like a resistor. A series of two junction behaves as two resistors in series. The net resistance is the sum of the two resistances. In the present problem, the resistance for each connection are $r_{L.R} = 1/(e^2\xi_{L.R})$. The above formula shows that the net resistance of the two-step tunneling

is the same as adding the two resistances. This result is expected once the energy levels in the QD are treated as continuous, and the charging energy U is neglected.

The Coulomb blockade is obtained by including the charging energy. The example is simplified by assuming that the two electrodes are identical, so $\xi_L = \xi_R$. Furthermore, it is assumed that the QD is exactly midway between the two electrodes. Then one must choose $V_{QD} = V/2$ by symmetry. For each value of V the ratio (8.394) is solved for the relevent number of n-values. Define this ratio as $r(2n+1)$. By normalizing all occupation numbers to f_0, we find for the partition function and the current

$$Z_{QD} = f_0 \left\{ 1 + r(1)[1 + r(3)(1 + \cdots)] + \frac{1}{r(-1)}\left[1 + \frac{1}{r(-3)}(1 + \cdots)\right] \right. \tag{8.401}$$

$$I = \frac{f_0}{Z_{QD}} \{R_L(1,0) - R_L(-1,0) + r(1)[R_L(2,1) - R_L(0,1)] + \frac{1}{r(-1)}[R_L(0,-1) - R_L(-2,-1)] + \cdots \} \tag{8.402}$$

A typical result from this formula is shown in Fig. 8.15. The current is an asymmetric function of the voltage. The onset of current begins at $e|V| = 2U$ since the current can only occur if $e|V|/2 > U$. Electrons can only hop from either electrode on to the QD if the voltage difference supplies the charging energy U. Since the QD is in the middle of the space between the electrodes, a voltage of V gives only $V/2$ between the QD and either electrode. Note that there is a second weak threshold when $e|V|/2 = 3U$ at $e|V| = 0.6\,\text{eV}$. This point is the threshold for having two extra electrons on the QD: the factor of three is $2^2 - 1^2$.

Since the geometry is perfectly symmetric, the current can also proceed by the generation of holes ($n < 0$). In this case, the first step is for an electron to hop from the QD to

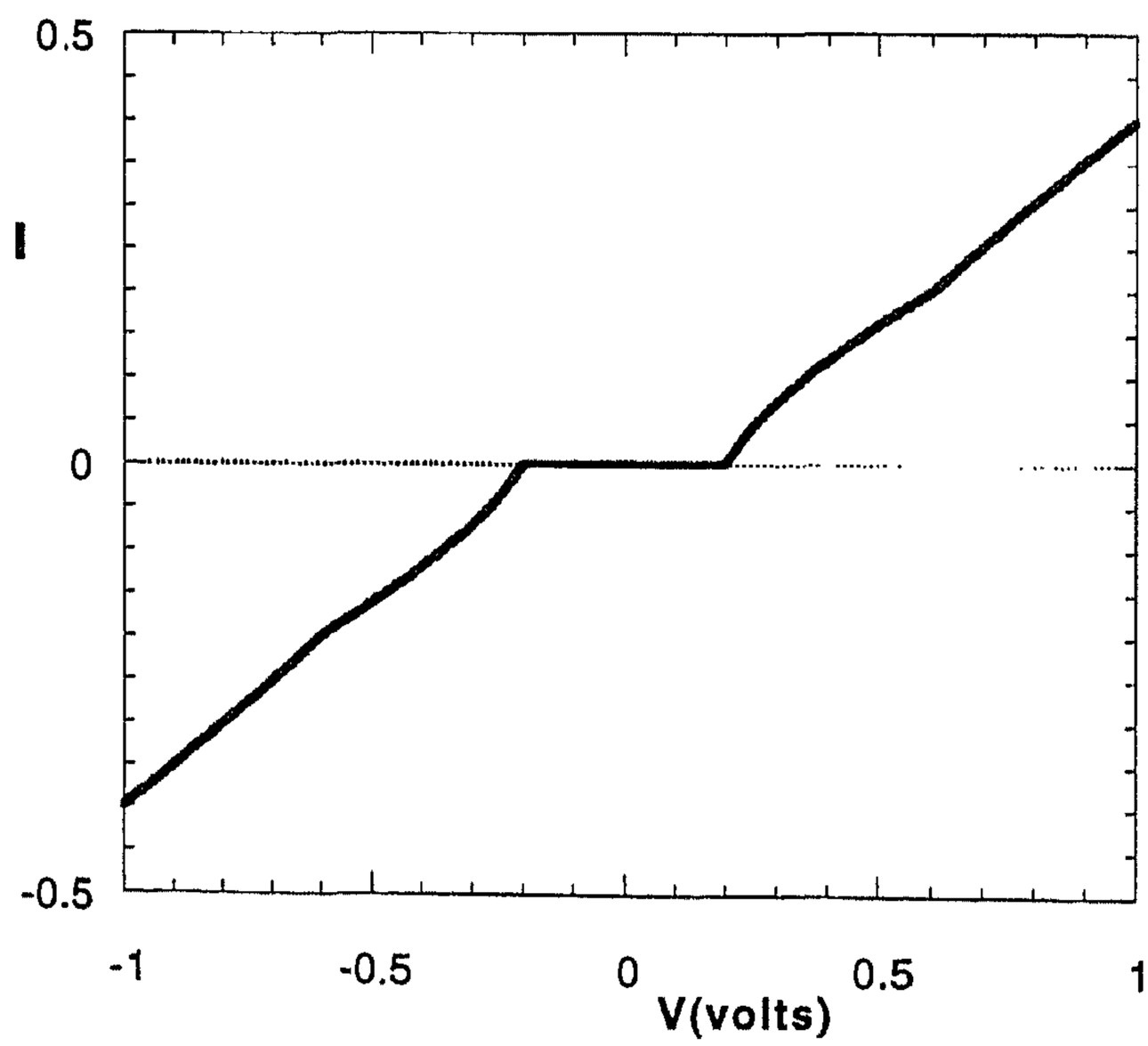


FIGURE 8.15 Current–voltage characteristic for a quantum dot placed symmetrically between two electrodes. $U = 0.1\,\text{eV}$ and $T = 10\,\text{K}$.

the left electrode, making a hole on the QD. The second step is for an electron to tunnel from the right electrode to the QD, which fills this hole state.

This example is typical for the case that the QD is large enough that its electron states can be considered to be continuous. A much different case, for a small QD, is when the energy levels are large compared to the thermal energy. Then the conductance is a series of steps, where each step is another available energy state on the QD.

An interesting experiment has a magnetic impurity in the QD (Goldhaber-Gordon *et al.*, 1998). The tunneling becomes a type of spectroscopy which can measure the density of states of the magnetic impurity. The experiment allows a test of solutions to the Anderson model.

8.6.4. Quantum Conductance

Ballistic transport is the case where an electron carries current without any scattering (Landauer, 1981, 1989). There are several relevent length scales: the length L of the sample, and the mean-free-path λ of the electron. Earlier in this chapter it was shown there are separate values of the mfp for: momentum scattering, energy scattering, electron–electron interactions, etc. Here the most relevant mfp is that for momentum scattering (λ_t), since it enters into the definition of the electrical conductivity. If $L < \lambda_t$ then the current is carried across the material without much chance of scattering. In this case the transport is called *ballistic*.

A special formula is needed for the current from ballistic transport. One cannot employ the Boltzmann equation or the Kubo formula, since they all assume a system with quasi-equilibrium, which means that the mfp is smaller than the dimensions of the sample. Instead, the correct picture is from simple quantum mechanics. Consider a particle moving to the right in one dimension, with a wave vector k. The number of particles per unit distance with wave vector k is

$$\frac{dk}{2\pi} n_F(\xi_k) \tag{8.403}$$

The Fermi–Dirac occupation number is for the source of electron on the left electrode. Assume there is another electrode on the right which is at a voltage V with respect to the one on the left. Then the number of electrons going from right to left is given by

$$\frac{dk}{2\pi} n_F(\xi_k + eV) \tag{8.404}$$

The net current is obtained by subtracting these two formula, and multiplying by ev_k, where v_k is the velocity

$$I = \frac{e}{2\pi} \int dk v_k [n_f(\xi_k) - n_F(\xi_k + eV)] \tag{8.405}$$

$$= \frac{e}{2\pi\hbar} \int d\xi_k [n_f(\xi_k) - n_F(\xi_k + eV)] \tag{8.406}$$

$$I = \sigma_0 V \tag{8.407}$$

$$\sigma_0 = \frac{e^2}{h} \tag{8.408}$$

The quantity e^2/h is the quantized conductance. Its units are interesting. In cgs units, $e^2/\hbar c$ is dimensionless, so that e^2/h has the units of velocity. That is not the correct result here.

Instead, in SI units e has the units of Coulombs, and Planck's constant has the units of Joule-second, which gives that σ_0 has the units of Siemens, which is the inverse of the Ohm. The approximate value is $\sigma_0 = 38.8\,\mu S$.

The above derivation applies to a single channel of electrons. Consider the conduction down a quantum wire. Treat the wire as a type of wave guide, which has many conducting modes, although each new one occurs at higher energy. The number of conducting modes as a function of energy is defined as $N_c(E)$. This quantity is a series of step functions as each new mode opens up along the quantum wire. Then the current is actually given by the expression

$$I = 2\sigma_0 N_c(eV)V \tag{8.409}$$

The factor of two in front is for spin degeneracy. Each orbital mode usually has spin degeneracy. This symmetry is broken in a magnetic field. Some writers prefer to define $2\sigma_0$ as the quantum of conductance.

PROBLEMS

1. Use the piezoelectric electron–phonon interaction in Sec. 1.3 to calculate the temperature dependence of the electrical conductivity in a semiconductor.

2. Consider the self-energy of an electron from unscreened exchange,

$$\Sigma(\mathbf{p}) = \frac{1}{\beta\nu}\sum_{\mathbf{q}} v_q \sum_{ip} \mathscr{G}(\mathbf{p}+\mathbf{q}, ip) \tag{8.410}$$

which was evaluated in Sec. 5.1.6. Derive the Ward identity for this self-energy—the equivalent of (8.81). Use this result to show that the Coulomb ladder diagrams have negligible effect upon the basic polarization $P^{(1)}(\mathbf{q}, i\omega)$ in the limit where $\mathbf{q} \to 0$.

3. The correlation function

$$\chi(\mathbf{q}, i\omega) = -\int_0^\beta d\tau e^{i\omega\tau}\langle T_\tau \rho(\mathbf{q},\tau)\rho(-\mathbf{q},0)\rangle \tag{8.411}$$

vanishes when $\mathbf{q} \to 0$. Show that the bare bubble $P^{(1)}(\mathbf{q}, i\omega)$ has this feature. Use the Ward identity to show that it still vanishes when self-energy functions and vertex functions are included in the evaluation.

4. Derive the rate of temperature relaxation in a semiconductor due to the deformation potential scattering of electrons by acoustical phonons. Give the answer for high and low temperatures. Evaluate the numerical value of the lifetimes using the data in Table 8.1.

5. Use polar coupling to optical phonons to calculate the rates of relaxation for an electron in a semiconductor: (a) for momentum, (b) for energy, and (c) temperature.

6. Use deformation coupling to optical phonons to calculate the rates of relaxation for an electron in a semiconductor: (a) for momentum, (b) for energy, and (c) for temperature.

7. Use (8.228) to evaluate $\gamma(\varepsilon + i\delta, \varepsilon + i\delta)$, and show that it equals unity, in agreement with the Ward identity.

8. Derive the Ward identity (8.219) for the phonon ladder diagrams.

9. Solve the scalar vertex equation (8.219) using the same techniques used to solve $\gamma(\mathbf{p}, \varepsilon + i\delta, \varepsilon + i\delta)$. Show that $\Gamma(\varepsilon + i\delta, \varepsilon + i\delta) = 1 - \partial\Sigma/\partial\varepsilon$ but that $\Gamma(\varepsilon + i\delta, \varepsilon - i\delta)$ does not obey the Ward identity. Furthermore, show for every solution $R(\varepsilon) = \Gamma(\varepsilon + i\delta, \varepsilon - i\delta)$ to its vertex equation that $R(\varepsilon) + a\ \mathrm{Im}[\Sigma(\varepsilon)]$ is also a solution, where a is an arbitrary constant.

10. Write down the collision rate $(\partial f/\partial t)$ for electron–electron interactions. Evaluate it for a semiconductor where the few conduction electrons obey Maxwell–Boltzmann statistics. Use Debye screening for the dielectric function.

11. Show that the leading term in the thermopower for a free-electron gas is

$$S = -\frac{\pi^2 k_B^2 T}{2eE_F} \tag{8.412}$$

12. Evaluate the correlation function $L^{(22)}$ for the thermal conductivity of a metal. Use the free-electron heat current operator, and show that the integral diverges for the first term in the vertex summation.

13. Solve Eqns. (8.215) and (8.235) in the limit of small temperature and show that the resistivity $\rho \sim T^5$ from phonons. Assume that $\alpha_{tr}^2 F(u) = gu^2$ at small u, where g is a constant.

14. Compare the relaxation time for transport deduced from the force-balance theory (8.198) with that given by (8.236). How do they differ? Show they become identical in the limit of high T.

15. Write down the rate equations for current between two electrodes separated by two identical quantum dots in series. What is the minimum voltage for current to flow at zero temperature?

Chapter 9

Optical Properties of Solids

9.1. NEARLY FREE-ELECTRON SYSTEMS

9.1.1. General Properties

This section is concerned with the optical properties of free-particle systems, such as electrons in metals or semiconductors. A completely free-electron system does not absorb light at all, so that its optical properties are uninteresting. The ability of the nearly free-particle system to absorb light is due to its imperfections or deviations from homogeneity. If these effects are small, then so is the light absorption. This situation is described here. The optical properties are calculated for simple metals and semiconductors. The best way to do this is using the so called *force–force correlation function.*

First it is useful to recall some of the fundamental formulas regarding the optical properties of solids. As shown in Sec. 4.5.3, the transverse dielectric function of the solid in the limit of long wavelength is

$$\varepsilon_{\perp}(\omega) = 1 - \frac{\omega_p^2}{\omega^2} - \frac{4\pi}{\omega^2}\pi(\omega) \tag{9.1}$$

$$\pi(i\omega) = -\frac{1}{V}\int_0^{\beta} d\tau e^{i\omega_n\tau}\langle T_\tau j_\mu(\tau) j_\mu(0)\rangle \tag{9.2}$$

where $\pi(\omega)$ is the retarded form of the current–current correlation function $\pi(i\omega)$. The optical wave vector $\mathbf{k}$ is so short that the correlation function is evaluated in the limit where $\mathbf{k} \to 0$. The results for $\mathbf{k} \to 0$ are sometimes different from those for $\mathbf{k} = 0$, in which case the limit result is appropriate. The $\mathbf{k}$ symbol is omitted in the notation. The system is assumed to be isotropic. The index μ is any direction (x, y, z), since the average value of $\langle j_x j_x\rangle$ is equal to that of $\langle \mathbf{j} \cdot \mathbf{j}\rangle/3$. The conductivity $\sigma(\omega)$ of the solid is given by

$$\varepsilon_{\perp}(\omega) = 1 + \frac{4\pi i\sigma(\omega)}{\omega} \tag{9.3}$$

$$\sigma(\omega) = \frac{ine^2}{m\omega} + \frac{i\pi(\omega)}{\omega} \tag{9.4}$$

with the usual identity $\mathrm{Re}(\sigma) = -\mathrm{Im}[\pi(\omega/\omega]$. The relationships are also valid for the real part of $\pi(\omega)$ which contributes to the imaginary part of the conductivity. These general formulas will be the basis for most of our discussion.

A general theorem which can be proved is the *f-sum rule*,

$$\varepsilon_\perp(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega) \tag{9.5}$$

$$\int_0^\infty \omega d\omega \varepsilon_2(\omega) = \frac{\pi\omega_p^2}{2} \tag{9.6}$$

The similar theorem for the longitudinal dielectric function was shown in Sec. 5.7. The theorem for the transverse dielectric function can also be written in terms of the conductivity since $\varepsilon_2 = 4\pi\,\mathrm{Re}(\sigma/\omega)$:

$$\int_0^\infty d\omega\,\mathrm{Re}[\sigma(\omega)] = \frac{\pi n_0 e^2}{2m} \tag{9.7}$$

The real part of the conductivity is the absorbing power of the solid. This theorem states that the absorbing power depends only on the density of particles n_0 and not on the detailed form of the interactions. Many-body theory can rearrange the absorbing power by moving it from one part of the spectrum to another, but it cannot alter the net amount. The statement in the opening paragraph, that the light absorption is small if the perturbations are small, needs clarification. Insight is provided by examining the Drude formula.

The Drude formula is an empirical relationship which is found to be valid, over a restricted frequency range, in most nearly free-electron systems. It states that

$$\pi(\omega) = -\frac{n_0 e^2}{m}\frac{1}{1 - i\omega\tau_0} \tag{9.8}$$

$$\mathrm{Re}[\sigma(\omega)] = \frac{n_0 e^2 \tau_0}{m(1 + \omega^2\tau_0^2)} \tag{9.9}$$

The quantity τ_0 is a relaxation time, which is usually taken to be a constant. Assume that the Drude formula is valid for all frequencies and with a constant relaxation time. Then the *f*-sum rule is satisfied independently of the size of the relaxation time:

$$\int_0^\infty d\omega\,\mathrm{Re}[\sigma(\omega)] = \frac{n_0 e^2 \tau_0}{m}\int_0^\infty \frac{d\omega}{1 + (\omega\tau_0)^2} = \frac{\pi n_0 e^2}{2m} \tag{9.10}$$

$\mathrm{Re}[\sigma(\omega)]$ is a Lorentzian distribution, whose width is $1/\tau_0$ and whose height is τ_0, so that the area under the curve is independent of τ_0.

The discussion of dc conductivity showed that inverse relaxation times are proportional to scattering rates. The purely free-particle system, without any inhomogeneity, corresponds to the case where $1/\tau_0 = 0$ or $\mathrm{Re}[\sigma(\omega)] \propto \delta(\omega)$. The scattering processes contribute to τ_0. If the Drude expression were rigorous, one would only need to find τ_0, which in fact is finding $1/\tau_0$. The limit $\omega \to 0$ in (9.9) gives the dc conductivity $\mathrm{Re}(\sigma) = n_0 e^2 \tau_0/m$.

If the scattering rate $1/\tau_0$ is small, then τ_0 is large, and the dimensionless product $\omega\tau_0$ is large except at very small frequencies. Expand (9.9) for $\omega\tau_0 \gg 1$ and find

$$4\pi\,\mathrm{Re}[\sigma(\omega)] = \frac{\omega_p^2}{\omega^2\tau_0}\left[1 + O\left(\frac{1}{(\omega\tau_0)^2}\right)\right] \tag{9.11}$$

At higher frequencies the real part of the conductivity is proportional to the scattering rate $1/\tau_0$, which is small by our assumption. This result is what is meant by the limit of weak scattering. The method of force–force correlation function casts $\mathrm{Re}[\sigma(\omega)]$ in this form, as an expansion in the scattering rate. The resulting theories are valid at high frequencies where $\omega\tau_0 \gg 1$.

9.1.2. Force–Force Correlation Functions

The first theorem to be proved has already been mentioned, namely that the homogeneous electron gas does not absorb any light. This assertion is proved by showing that the Hamiltonian commutes with the momentum operator $\mathbf{P} = \sum \mathbf{k} C^\dagger_{\mathbf{k},\sigma} C_{\mathbf{k}\sigma}$

$$H_0 = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \frac{1}{2\nu} \sum_{\mathbf{pkq}} \sum_{\sigma\sigma'} v_q C^\dagger_{\mathbf{k}+\mathbf{q},\sigma'} C^\dagger_{\mathbf{p}-\mathbf{q},\sigma} C_{\mathbf{p}\sigma} C_{\mathbf{k}\sigma'} \tag{9.12}$$

$$0 = \frac{d}{dt}\mathbf{P} = i[H_0, \mathbf{P}] \tag{9.13}$$

The momentum operator has no time dependence and is a constant of motion.

The absorption of light is described by the Kubo formula, which is the correlation of the momentum operator with itself:

$$\mathrm{Re}[\sigma(\omega)] = \frac{e^2}{6m^2\nu}\left(\frac{1 - e^{-\beta\omega}}{\omega}\right)\int_{-\infty}^{\infty} dt e^{i\omega t} \langle \mathbf{P}(t) \cdot \mathbf{P}(0)\rangle \tag{9.14}$$

where the current is proportional to the momentum $\mathbf{j}(t) = (e/m)\mathbf{P}(t)$. The lack of time variation in $\mathbf{P}$ means that the real part of the conductivity is identically zero for any nonzero frequency. There is no light absorption for $\omega \neq 0$.

A simple way of deriving the force–force correlation function was given by Hopfield (1965). The argument proceeds in two steps. The first step is to replace $\mathbf{P}(t)$ by the identical operator $(i/\omega)d\mathbf{P}/dt$, which is justified on the grounds that in an ac field of frequency ω the time dependence of interest is just $\exp(-i\omega t)$. This replacement is done for both $\mathbf{P}$ operators, which gives the expression

$$\mathrm{Re}[\sigma(\omega)] = \frac{e^2}{6m^2\omega^3\nu}(1 - e^{-\beta\omega})\int_{-\infty}^{\infty} dt e^{i\omega t}\left\langle \frac{d\mathbf{P}(t)}{dt} \cdot \left[\frac{d\mathbf{P}(t')}{dt'}\right]_{t'=0}\right\rangle \tag{9.15}$$

The second step in the argument is to notice that the time derivative of the momentum is just the force $\mathbf{F}(t) = d\mathbf{P}/dt$, so the force–force correlation function is

$$\mathrm{Re}[\sigma(\omega)] = \frac{e^2}{6m^2\omega^3\nu}(1 - e^{-\beta\omega})\int_{-\infty}^{\infty} dt e^{i\omega t} \langle \mathbf{F}(t) \cdot \mathbf{F}(0)\rangle \tag{9.16}$$

This derivation is essentially correct and will be justified later with more rigor. The force–force correlation function then serves as another possible starting point for the evaluation of the conductivity. For free-particle systems, it usually leads to a usable answer with much less work than the current–current correlation function, because the first term in the perturbation expansion of the force–force correlation function is usually the answer. For example, if the force is due to phonons, then $F \propto M_\lambda(\mathbf{q})$, where $M_\lambda(\mathbf{q})$ is the electron–phonon matrix element. The leading term in the correlation function is proportional to M_λ^2, which is often the largest term. The remaining parts of the correlation function are evaluated assuming the particles are free, which is quite easy. Similarly, if the absorption is due to impurities, the

leading term is proportional to concentration n_i of impurities, which is often the important term. The advantage of the force–force correlation function is that the leading term is often adequate.

One calculation for which the force–force correlation function would be a poor starting point would be the dc conductivity. The correlation function has a prefactor of ω^{-3}, and obviously a great deal of cancellation in the correlation function has to take place in the limit $\omega \to 0$ in order to get $\sigma = n_0 e^2 \tau_0 / m$.

A rigorous derivation of the force–force correlation function was given independently by Mahan (1970) and Hasegawa and Watabe (1969). The trick is to start from the Kubo formula and integrate by parts on the τ variable:

$$\tau(i\omega) = -\frac{1}{\nu}\int_0^\beta d\tau e^{i\omega_n \tau}\langle T_\tau j_\mu(\tau) j_\mu(0)\rangle \tag{9.17}$$

$$\int u dv = uv - \int v du \tag{9.18}$$

$$u = j_\mu(\tau), \qquad v = \frac{e^{i\omega\tau}}{i\omega} \tag{9.19}$$

The integration by parts gives

$$\pi(i\omega) = -\frac{1}{i\omega\nu}[\langle j_\mu(\beta) j_\mu(0) - j_\mu(0) j_\mu(0)]\rangle + \frac{1}{i\omega\nu}\int_0^\beta d\tau e^{i\omega_n \tau}\left\langle T_\tau \frac{dj_\mu(\tau)}{d\tau} j_\mu(0)\right\rangle \tag{9.20}$$

The first term is the constant of integration. These constants have the general form of commutators:

$$\langle A(\beta)B(0) - A(0)B(0)\rangle = \mathrm{Tr}(Ae^{-\beta H}B - e^{-\beta H}AB) \tag{9.21}$$

$$= \mathrm{Tr}[e^{-\beta H}(BA - AB)] = -\langle [A, B]\rangle \tag{9.22}$$

where the cyclic properties of the trace are used to rearrange the first term. The one in (9.20) is zero since it is the commutator of $j_\mu(0)$ with itself.

The next step in the derivation is to take the integration by parts on the other current operator $j_\mu(0)$. Since the correlation function depends only on the difference of the two τ values in its argument, this correlation function is also equal to $\langle \mathrm{Tr}[(dj_\mu(\tau')/(d\tau')_{\tau'=0} j_\mu(-\tau)\rangle$. The integration by parts on the term $j_\mu(-\tau)$ gives

$$\pi(i\omega) = \frac{1}{(i\omega)^2\nu}\left\langle\left[j_\mu(0), \left[\frac{dj_\mu(\tau)}{d\tau}\right]_{\tau=0}\right]\right\rangle$$
$$-\frac{1}{(i\omega)^2\nu}\int_0^\beta d\tau e^{i\omega_n \tau}\left\langle T_\tau\left[\frac{dj_\mu(\tau)}{d\tau}\right]\left[\frac{dj_\mu(\tau')}{d\tau'}\right]_{\tau'=0}\right\rangle \tag{9.23}$$

The first term contains the commutator of j_μ with $dj_\mu/d\tau$. This commutator yields a constant, which is real and so does not contribute to the real part of the conductivity or to the absorption of light. But Simanek (1971) has shown that this term does contribute to the renormalization of the optical mass so it is retained. The other term in (9.23) is the frequency–dependent part, which is indeed the correlation function of $dj/d\tau$ with itself. The current is just proportional to the momentum $\mathbf{j} = (e/m)\mathbf{P}$, so (9.15) is derived again.

The Hamiltonian is written as $H = H_0 + V$, where H_0 is the homogeneous electron gas in (9.12), while V is the potential which causes other forces besides electron–electron inter-

actions. For example, the potential could be the sum of the interactions with the crystalline potential $V_{\mathbf{G}}$ where the $\mathbf{G}$ are reciprocal lattice vectors, with impurities of density operator $\rho_i(\mathbf{q})$ (see Sec. 4.1.5), or with phonons:

$$V = \sum_{\mathbf{q}} \rho(\mathbf{q})\Phi(\mathbf{q}) \tag{9.24}$$

$$\Phi(\mathbf{q}) = \delta_{\mathbf{q}=\mathbf{G}} V_{\mathbf{G}} + \frac{V_i(\mathbf{q})}{\nu}\rho_i(\mathbf{q}) + \sum_{\lambda} \frac{M_\lambda(\mathbf{q})}{\sqrt{\nu}}(a_{\mathbf{q}\lambda} + a^\dagger_{-\mathbf{q}\lambda}) \tag{9.25}$$

The time derivative of the current operator is

$$\frac{dj_\mu}{d\tau} = [H, j_\mu] = [V, j_\mu] = -\sum_{\mathbf{q}} \Phi(\mathbf{q})[j_\mu, \rho(\mathbf{q})] \tag{9.26}$$

$$[j_\mu, \rho(\mathbf{q})] = \frac{e}{m}\sum_{\mathbf{k}\mathbf{p}ss'} k_\mu [C^\dagger_{\mathbf{k}s} C_{\mathbf{k}s}, C^\dagger_{\mathbf{p}+\mathbf{q},s'} C_{\mathbf{p},s'}] \tag{9.27}$$

$$= \frac{e}{m}\sum_{\mathbf{k}s} C^\dagger_{\mathbf{k}+\mathbf{q},s} C_{\mathbf{k},s}[(k_\mu + q_\mu) - k_\mu] = \frac{e}{m} q_\mu \rho(\mathbf{q}) \tag{9.28}$$

$$\frac{dj_\mu}{d\tau} = -\frac{e}{m}\sum_{\mathbf{q}} q_\mu \rho(\mathbf{q})\Phi(\mathbf{q}) \tag{9.29}$$

The last commutator is just Newton's law: the time derivative of the momentum is the force, which is the gradient of the potential

$$F_\mu(\mathbf{r}) = -\nabla_\mu V(\mathbf{r}) = -i\sum_{\mathbf{q}} V(\mathbf{q}) q_\mu e^{i\mathbf{q}\cdot\mathbf{r}} \tag{9.30}$$

The evaluation of the constant term in (9.23) requires the additional commutator

$$\left[j_\mu, \frac{dj_\mu}{d\tau}\right] = -\frac{e}{m}\sum_{\mathbf{q}} q_\mu \Phi(\mathbf{q})[j_\mu, \rho(\mathbf{q})] \tag{9.31}$$

$$= -\frac{e^2}{m^2}\sum_{\mathbf{q}} q_\mu^2 \rho(\mathbf{q})\Phi(\mathbf{q}) \tag{9.32}$$

Using these results in (9.23) gives

$$\pi(i\omega) = \frac{e^2}{\nu m^2 (i\omega)^2}\left[-\sum_{\mathbf{q}} q_\mu^2 \langle \rho(\mathbf{q})\Phi(\mathbf{q})(\mathbf{q})\rangle + \sum_{\mathbf{q}\mathbf{q}'} q_\mu q'_\mu \int_0^\beta d\tau e^{i\omega_n \tau} \right.$$
$$\left. \times \langle T_\tau \rho(\mathbf{q}, \tau)\Phi(\mathbf{q}, \tau)\rho(\mathbf{q}', 0)\Phi(\mathbf{q}', 0)\rangle\right] \tag{9.33}$$

The generalized potential $\Phi(\mathbf{q})$ is given a τ dependence, which applies only to the phonon part of the three terms in (9.25). The second term is the force–force correlation function, which is equivalent to (9.16). It is often sufficient to evaluate this expression only to order Φ^2. The second term is already evaluated to this order, so that the remaining correlation function is

evaluated ignoring the interaction V. In this case $\mathbf{q}' = -\mathbf{q}$. The first term in (9.33) must also be evaluated to order Φ^2 by expanding the S matrix for the interaction potential V:

$$\begin{aligned}\langle\rho(\mathbf{q})\Phi(\mathbf{q})\rangle &= \mathrm{Tr}[e^{-\beta H_0}S(\beta)\rho(\mathbf{q})\Phi(\mathbf{q})] \\ &= \mathrm{Tr}[e^{-\beta H_0}\rho(\mathbf{q})\Phi(\mathbf{q})] - \int_0^\beta d\tau\langle T_\tau V(\tau)\rho(\mathbf{q},0)\Phi(\mathbf{q},0)\rangle + O(\Phi^3) \qquad (9.34)\end{aligned}$$

Only the term in V is retained in order to have a result proportional to Φ^2. The first term on the right ${}_0\langle\rho(\mathbf{q})\Phi(\mathbf{q})\rangle_0$ is zero. Combining both terms in (9.33) gives

$$\pi(i\omega) = -\frac{e^2}{\nu m^2(i\omega)^2}\left[\sum_{\mathbf{q}} q_\mu^2\int_0^\beta d\tau(e^{i\omega\tau}-1)\langle T_\tau\rho(\mathbf{q},\tau)\rho(-\mathbf{q},0)\rangle\langle T_\tau\Phi(\mathbf{q},\tau)\Phi(-\mathbf{q},0)\rangle\right] \qquad (9.35)$$

The term with $\exp(i\omega\tau)$ comes from the force–force correlation function, while the -1 term comes from the constant term.

This formula will be evaluated for several different types of potentials in the remaining parts of this section. The type of formula which results can be illustrated by a simple example. Take the case where the potential Φ is independent of τ. One example is the scattering from impurities, where, according to Sec. 4.1.5,

$$\langle\Phi(\mathbf{q})\Phi(\mathbf{q}')\rangle = \frac{1}{\nu^2}\langle V_i(\mathbf{q})V_i(\mathbf{q}')\langle\rho_i(\mathbf{q})\rho_i(\mathbf{q}')\rangle \qquad (9.36)$$

$$= \frac{n_i}{\nu}V_i(\mathbf{q})^2\delta_{\mathbf{q}+\mathbf{q}'=0} \qquad (9.37)$$

where $V_i(\mathbf{q})$ is the unscreened potential between the electron and the impurity. The remaining part of the τ integral in (9.35) is the definition of the longitudinal dielectric function from Sec. 5.4:

$$-\frac{1}{\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau\rho(\mathbf{q},\tau)\rho(-\mathbf{q},0)\rangle = \frac{q^2}{4\pi e^2}\left[\frac{1}{\varepsilon(\mathbf{q},i\omega)}-1\right] \qquad (9.38)$$

$$\pi_i(i\omega) = \frac{n_i}{4\pi m^2\nu(i\omega)^2}\sum_{\mathbf{q}} q_\mu^2 q^2 V_i(\mathbf{q})^2\left[\frac{1}{\varepsilon(\mathbf{q},i\omega)}-\frac{1}{\varepsilon(\mathbf{q})}\right] \qquad (9.39)$$

The subscript i denotes the impurity contribution to $\pi(i\omega)$. The current–current correlation function is expressed as the difference between $1/\varepsilon(\mathbf{q},i\omega)$ and $1/\varepsilon(\mathbf{q},0)$. Since $\varepsilon(\mathbf{q},i\omega)$ is an even function of $i\omega$, in the limit where $i\omega \to 0$ it must behave as $\varepsilon(\mathbf{q},i\omega) = \varepsilon(\mathbf{q},0) + O(i\omega)^2$, which shows that π_i does not diverge in this limit. The retarded function is found by $i\omega \to \omega + i\delta$:

$$\pi_i(\omega) = \frac{n_i}{4\pi m^2\nu\omega^2}\sum_{\mathbf{q}} q_\mu^2 q^2 V_i(\mathbf{q})^2\left[\frac{1}{\varepsilon(\mathbf{q},\omega)}-\frac{1}{\varepsilon(\mathbf{q})}\right] \qquad (9.40)$$

$$\mathrm{Re}[\sigma_i(\omega)] = -\frac{n_i}{4\pi m^2\omega^3}\int\frac{d^3q}{(2\pi)^3}q_\mu^2 q^2 V_i(\mathbf{q})^2\,\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q},\omega)}\right] + O(V_i^3) \qquad (9.41)$$

Equation (9.41) is the Hopfield (1965) formula, which shows that the real part of the transverse conductivity is proportional to the imaginary part of the inverse longitudinal dielectric function. The latter quantity is interpreted as the rate of making excitations in the electron gas. The light absorption occurs because electronic excitations are created, while the creation rate must depend on the concentration of impurities n_i in the system. This inter-

pretation agrees with our original assertion that the electron gas can absorb light only when the impurities, or other inhomogeneities, are present to dissipate the momentum. An obvious advantage of the Hopfield formula is that the complicated aspects of electron–electron interactions are naturally included in the dielectric function, for which one can use the model forms described in Chapter 5.

9.1.3. Fröhlich Polarons

The optical absorption of free polarons was first calculated by Gurevich *et al.* (1962), who started from the usual form of the Kubo formula. In the Fröhlich Hamiltonian (7.1),

$$H = \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \omega_0 \sum_{\mathbf{q}} a^\dagger_{\mathbf{q}} a_{\mathbf{q}} + \frac{M_0}{\sqrt{\nu}} \sum_{\mathbf{q}} \frac{1}{|\mathbf{q}|} C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}\sigma} (a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}})$$
$$M_0^2 = \frac{4\pi\alpha\omega_0^{3/2}}{\sqrt{2m}} \tag{9.42}$$

the electron–phonon interaction V is expanded using the usual S-matrix techniques. The real part of the conductivity is calculated to order $O(\alpha)$, which is the first nonvanishing term. The three diagrams which enter proportional to α are shown in Fig. 9.1. The first two are self-energy diagrams, while the third is a vertex contribution. The final result is obtained by adding contributions from all three diagrams. At zero temperature it is

$$\mathrm{Re}[\sigma(\omega)] = \frac{2e^2 n_0 \alpha \omega^{3/2}}{3m\omega^3} \sqrt{\omega - \omega_0}\, \Theta(\omega - \omega_0) + O(\alpha^2) \tag{9.43}$$

The free polaron can absorb light only when the frequency $\omega > \omega_0$. Polarons can emit LO phonons only when the polaron energy exceeds the phonon energy ω_0. The light wave excites the polaron to this energy, and then the polaron can emit the phonon. The light absorption happens by the emission of LO phonons, where the electron serves as the intermediary in the process. Of course, there is also a direct coupling between the photon and the TO phonons, which was discussed in Sec. 4.6 on polaritons.

Formula (9.43) can be derived from the correlation functions in (9.35). The part involving the potential is now just the Green's function for LO phonons:

$$\Phi(\mathbf{q}) = \frac{M_0}{q\sqrt{\nu}} (a_{\mathbf{q}} + a^\dagger_{-\mathbf{q}}) \tag{9.44}$$

$$\langle T_\tau \Phi(\mathbf{q}, \tau) \Phi(-\mathbf{q}, 0) \rangle = -\frac{M_0^2}{\nu q^2} \mathscr{D}^{(0)}(\tau) \tag{9.45}$$

FIGURE 9.1

Denote the density–density correlation function by the symbol $\chi(\mathbf{q}, \tau)$. The τ integral in (9.35) is evaluated by changing it into a summation over Matsubara frequencies, which is evaluated by the usual contour integral:

$$\chi(\mathbf{q}, \tau) = -\langle T_\tau \rho(\mathbf{q}, \tau)\rho(-\mathbf{q}, 0)\rangle \tag{9.46}$$

$$\begin{aligned}\pi(i\omega) &= -\frac{M_0^2 e^2}{m^2(i\omega)^2\nu}\sum_{\mathbf{q}}\frac{q_\mu^2}{q^2}\int_0^\beta d\tau(e^{i\omega\tau}-1)\chi(\mathbf{q},\tau)\mathscr{D}^{(0)}(\tau) + O(M_0^4)\\ &= -\frac{M_0^2 e^2}{m^2(i\omega)^2\nu}\sum_{\mathbf{q}}\frac{q_\mu^2}{q^2}\frac{1}{\beta}\sum_{iq}\mathscr{D}^{(0)}(iq)[\chi(\mathbf{q}, i\omega + iq) - \chi(\mathbf{q}, iq)]\\ &= -\frac{M_0^2 e^2}{m^2(i\omega)^2\nu}\sum_{\mathbf{q}}\frac{q_\mu^2}{q^2}\int_{-\infty}^{\infty}\frac{d\varepsilon}{2\pi}n_B(\varepsilon)\{2\,\mathrm{Im}[D^{(0)}(\varepsilon)]\chi(\mathbf{q}, i\omega+\varepsilon)\\ &\quad + 2\,\mathrm{Im}[\chi(\mathbf{q},\varepsilon)]\mathscr{D}^{(0)}(\varepsilon - i\omega) - 2\,\mathrm{Im}[D^{(0)}(\varepsilon)\chi(\mathbf{q},\varepsilon)]\}\end{aligned} \tag{9.47}$$

Take the retarded function $i\omega \to \omega + i\delta$, and the real part of the conductivity is the imaginary part of the preceding expression, where the last term does not contribute:

$$\mathrm{Re}[\sigma(\omega)] = \frac{2M_0^2 e^2}{m^2\omega^3}\int\frac{d^3q}{(2\pi)^3}\frac{q_\mu^2}{q^2}\int_{-\infty}^{\infty}\frac{d\varepsilon}{2\pi}n_B(\varepsilon) \tag{9.48}$$

$$\times\{\mathrm{Im}[D^{(0)}(\varepsilon)]\mathrm{Im}[\chi(\mathbf{q}, \varepsilon+\omega)] - \mathrm{Im}[\chi(\mathbf{q},\varepsilon)]\mathrm{Im}[D^{(0)}(\varepsilon-\omega)]\} \tag{9.49}$$

This result is general and applies to all electron–phonon systems. Sometimes it is useful to change variables $\varepsilon \to \varepsilon + \omega$ in the second term, which gives the equivalent expression

$$\mathrm{Re}\,\sigma(\omega) = \frac{2M_0^2 e^2}{m^2\omega^3}\int\frac{d^3q}{(2\pi)^3}\frac{q_\mu^2}{q^2}\int_{-\infty}^{\infty}\frac{d\varepsilon}{2\pi}[n_B(\varepsilon) - n_B(\varepsilon+\omega)] \tag{9.50}$$

$$\times\,\mathrm{Im}[D^{(0)}(\varepsilon)]\mathrm{Im}[\chi(\mathbf{q}, \varepsilon+\omega)] \tag{9.51}$$

The spectral function for the phonon Green's function is easy to evaluate:

$$-2\,\mathrm{Im}[D^{(0)}(\varepsilon)] = 2\pi[\delta(\varepsilon-\omega_0) - \delta(\varepsilon+\omega_0)] \tag{9.52}$$

The last step in the formal derivation is to determine $\chi(q, \omega)$. The free polarons are assumed to have small concentration n_0, so that electron–electron interactions can be neglected. Then the density–density correlation function is adequately approximated by the simple bubble diagram:

$$\chi(\mathbf{q}, i\omega) = P^{(1)}(\mathbf{q}, i\omega) = \frac{2}{\nu\beta}\sum_{\mathbf{p}, ip}\mathscr{G}^{(0)}(p)\mathscr{G}^{(0)}(p+q) \tag{9.53}$$

$$\begin{aligned}P^{(1)}(\mathbf{q},\omega) &= 2\int\frac{d^3p}{(2\pi)^3}n_p\left(\frac{1}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega + i\delta} + \frac{1}{\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} - \omega - i\delta}\right)\\ -2\,\mathrm{Im}[P^{(1)}(\mathbf{q},\omega)] &= 2\pi\int\frac{d^3p}{(2\pi)^3}n_p[\delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} + \omega) - \delta(\varepsilon_{\mathbf{p}} - \varepsilon_{\mathbf{p}+\mathbf{q}} - \omega)]\end{aligned} \tag{9.54}$$

The factor of 2 in front is for the spin degeneracy. The correction terms to this simple bubble term, from the electron–phonon interaction, contributes to higher-order terms in α.

The polarons are assumed to obey Maxwell–Boltzmann statistics, so the particle density $n_{\mathbf{p}}$ per spin state is

$$n_{\mathbf{p}} = \frac{n_0}{2}\left(\frac{2\pi\beta}{m}\right)^{3/2} \exp[-\beta\varepsilon_{\mathbf{p}}] \tag{9.55}$$

The present calculation is for zero temperature. The particles all approach the state with $\mathbf{p} \to 0$, so from (9.54) the spectral function is

$$-2\,\mathrm{Im}[\chi(\mathbf{q}, \omega)] = 2\pi n_0[\delta(\omega - \varepsilon_{\mathbf{q}}) - \delta(\omega + \varepsilon_{\mathbf{q}})] \tag{9.56}$$

Now evaluate the real part of the conductivity given in (9.51). Both spectral functions are given by delta functions, for phonons in (9.52) and for electrons in (9.56). One set of delta functions are used to eliminate the $d\varepsilon$ integral, which gives

$$\begin{aligned}\mathrm{Re}[\sigma(\omega)] = \pi\frac{2M_0^2e^2n_0}{m^2\omega^3}\int\frac{d^3q}{(2\pi)^3}\frac{q_\mu^2}{q^2}\{&[n_B(\omega_0) - n_B(\omega_0 + \omega)]\\ &\times[\delta(\omega_0 + \omega - \varepsilon_{\mathbf{q}}) - \delta(\omega_0 + \omega + \varepsilon_{\mathbf{q}})]\\ &+[1 + n_B(\omega_0) + n_B(\omega - \omega_0)][\delta(\omega - \omega_0 - \varepsilon_{\mathbf{q}}) - \delta(\omega - \omega_0 + \varepsilon_{\mathbf{q}})]\}\end{aligned} \tag{9.57}$$

At zero temperature, the boson occupation factor is $n_B(\omega) = -\Theta(-\omega)$. This expression simplifies for $\omega > 0$, $\cos^2\theta = q_\mu^2/q^2$ to

$$\mathrm{Re}[\sigma(\omega)] = \pi\frac{2M_0^2e^2n_0}{m^2\omega^3}\int\frac{d^3q}{(2\pi)^3}\cos^2\theta\,\delta(\omega - \omega_0 - \varepsilon_{\mathbf{q}}) \tag{9.58}$$

$$= \pi\frac{2M_0^2e^2n_0}{m^2\omega^3(2\pi)^3}\left(\frac{2\pi}{3}\right)(2m)^{3/2}\sqrt{\omega - \omega_0}\,\Theta(\omega - \omega_0) \tag{9.59}$$

Using the form for the electron–phonon matrix element M_0^2 gives the result (9.43) of Gurevich *et al.* The derivation has been remarkably short. The general form of the answer in (9.51), as the product of two spectral functions, is quite familiar from other calculations, and in this case both spectral functions are delta functions.

The optical absorption of Fröhlich polarons has been calculated extensively by Devreese (1972). He calculated the terms to order $O(\alpha^2)$, which are the two phonon terms. The polaron can be excited by the light and make two phonons as well as one. Results are shown in Fig. 9.2 for the value of the coupling constant $\alpha = 1$. The solid line is the result using (9.43) which includes just the one-phonon processes, while the dashed line is the result including both one- and two-phonon processes. The rather surprising result emerges that there is no sudden rise in the absorption at the two-phonon threshold $\omega = 2\omega_0$. The sharp rise at $\omega = \omega_0$ is not repeated as a sharp rise at $2\omega_0$. Instead, there is a very gradual rise starting at $2\omega_0$ and a slight decrease in the peak height. The two-phonon events make only a small change in the result. The polaron would rather emit two phonons as two separate events of one-phonon emission and not as a correlated event where two phonons are emitted. This conclusion is in accord with the remarks in Sec. 7.1 about the correlation between two phonons being small in the polaron cloud. Devreese (1972) has also calculated the optical absorption of polarons by the Feynman method of path integrals as well as by strong coupling theory. At larger values of polaron constant α he finds sharp absorption lines, which are interpreted as arising from internal states of the localized electron in the small polaron.

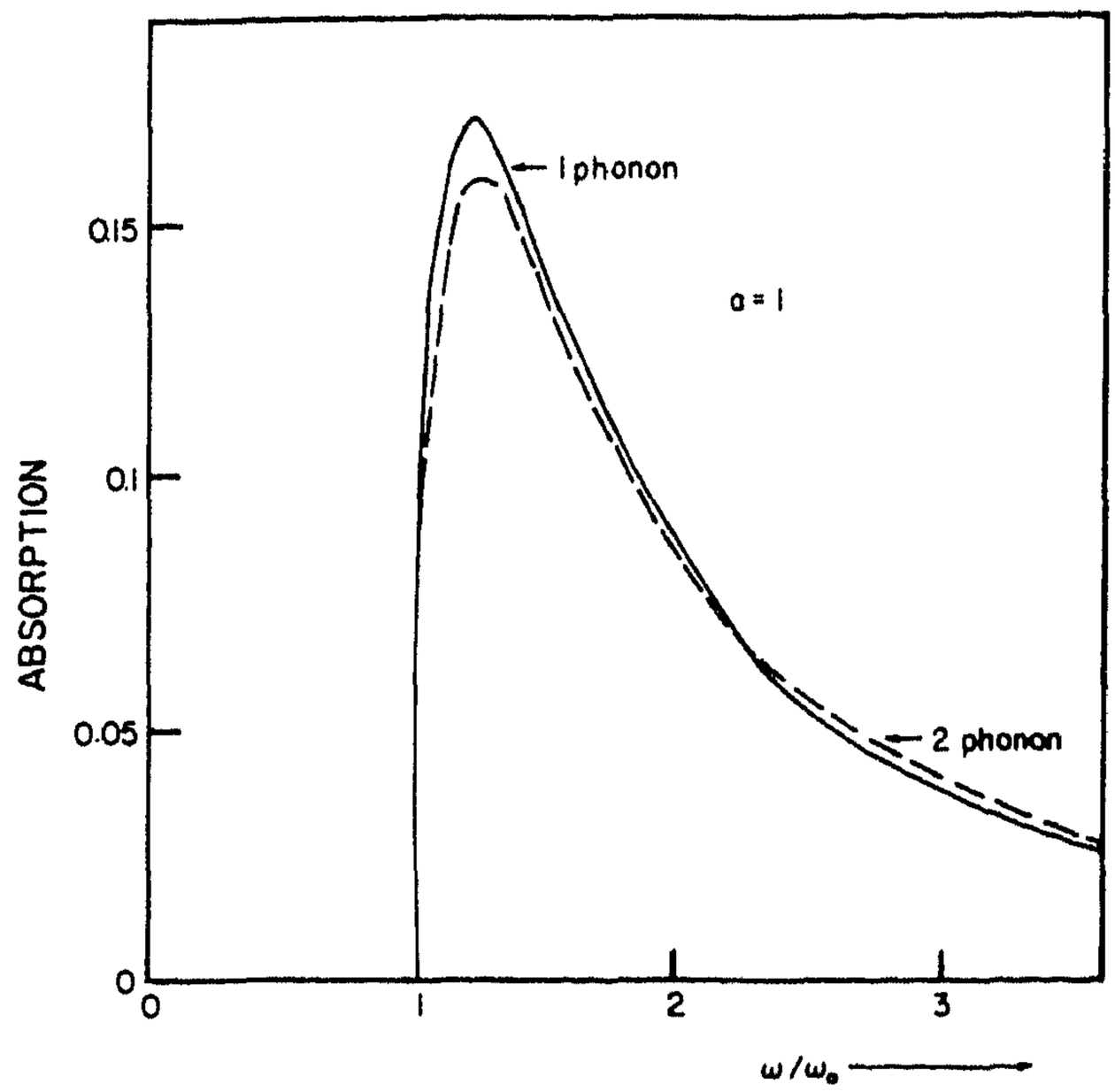


FIGURE 9.2 Theoretical prediction for the optical absorption of free polarons with $\alpha = 1$ and $T = 0\,\mathrm{K}$. The solid line is one-phonon theory, and the dashed line has two-phonon correlations. *Source*: Devreese (1972, p. 93).

9.1.4. Interband Transitions

Another application of the force–force correlation function is to calculate the rate of light absorption by interband transitions in a metal. Here the potential which provides the inhomogeneity is the crystalline potential of the atoms. This potential is periodic in crystalline solids, so it contributes momentum in discrete units called reciprocal lattice vectors **G**. This contribution was mentioned earlier as the first term in (9.25):

$$\Phi(\mathbf{q}) = \sum_{\mathbf{G}} V_{\mathbf{G}} \delta_{\mathbf{q}=\mathbf{G}} \tag{9.60}$$

This potential is fixed and has no correlation with itself: $\langle \Phi(\mathbf{q})\Phi(\mathbf{q}')\rangle = \langle \Phi(\mathbf{q})\rangle\langle \Phi(\mathbf{q}')\rangle$. However, the other correlation function in (9.35) is $\langle \rho(\mathbf{q}, \tau)\rho(\mathbf{q}', 0)\rangle$ and is nonzero only when $\mathbf{q} = -\mathbf{q}'$:

$$\langle \Phi(\mathbf{q})\Phi(-\mathbf{q})\rangle = \sum_{\mathbf{G}} V_{\mathbf{G}} V_{-\mathbf{G}} \delta_{\mathbf{q}=\mathbf{G}} = \sum_{\mathbf{G}} V_{\mathbf{G}}^2 \delta_{\mathbf{q}=\mathbf{G}} \tag{9.61}$$

This potential is static, so the correlation function of the potential is a constant. This case is similar to that of impurity scattering, which was given in (9.41). The impurity result can be used for the crystalline potential by changing $n_i V_i(\mathbf{q})^2$ to $V_{\mathbf{G}}^2 \delta_{\mathbf{q}=\mathbf{G}}$. The interband absorption is

$$\mathrm{Re}[\sigma_I(\omega)] = -\frac{1}{4\pi m^2 \omega^3} \sum_{\mathbf{G}} G_\mu^2 G^2 V_{\mathbf{G}}^2 \,\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{G}, \omega)}\right] + O(V_{\mathbf{G}}^3) \tag{9.62}$$

The subscript I stands for interband contribution. Note the following:

$$-\mathrm{Im}\left(\frac{1}{\varepsilon}\right) = \frac{\varepsilon_2}{|\varepsilon|^2} \tag{9.63}$$

where $1/|\varepsilon|^2$ is usually grouped with $V_{\mathbf{G}}^2$ to have a screened electron–ion interaction $|V_{\mathbf{G}}/\varepsilon(\mathbf{G}, \omega)|^2$. If the RPA result in Sec. 5.5 is used for $\varepsilon_2(\mathbf{G}, \omega)$, then when $G > 2k_F$,

$$\begin{aligned} \varepsilon_2(G, \omega) &= \frac{e^2 m}{G^3}\left[k_F^2 - \left(\frac{m}{G}\right)^2(\omega - \varepsilon_G)^2\right] \\ \mathrm{Re}[\sigma_l(\omega)] &= \frac{e^2}{4\pi m\omega^3}\sum_{\mathbf{G}}\frac{G_\mu^2}{G}\left|\frac{V_{\mathbf{G}}}{\varepsilon(G, \omega)}\right|^2\left[k_F^2 - \left(\frac{m}{G}\right)^2(\omega - \varepsilon_G)^2\right] \end{aligned} \tag{9.64}$$

This formula is nearly identical to a standard formula for describing interband transitions in metals: the Wilson–Butcher formula.

The Wilson–Butcher formula is used to describe the interband transition in the alkali metals; Figure 9.3(a) shows the band structure in an alkali metal in the $k = (110)$ direction. The Fermi level is shown as the dashed horizontal line. The interband transitions are shown, where the transition appears vertical in wave vector space since the wave vector of light is too small to show up as a horizontal deflection of the arrow. This transition is shown in the reduced zone scheme. Of course, in the extended zone description of Fig. 9.3(b), the optical transition really changes the wave vector of the electron by a reciprocal lattice vector $\mathbf{G}$. This change in the momentum of the electron, during the optical transition, is precisely what it needs to have the transition. The optical absorption occurs because the electron gains energy from the light wave and momentum from the rigid crystal lattice.

The optical absorption one expects from an alkali metal is shown in Fig. 9.4. At low frequency there is a Drude contribution, which comes from phonons. This contribution is well described by the Drude formula (9.9), where τ_0 is largely provided by the intrinsic scattering from phonons. The interband transition of Fig. 9.4 starts about $\hbar\omega = 2$ eV, with a linear rise at threshold. Since the Fermi surface does not touch the zone face, $k_F < G/2$, and (9.64) is the appropriate form to use for $\varepsilon_2(G, \omega)$. Figure 9.4 shows the experimental result of Smith (1969) for sodium, which has exactly this behavior. Similar interband terms were found by Smith (1970) for K, Rb, and Cs and in Li by Myers and Sixtensson (1976). The agreement between theory and experiment appears satisfactory for sodium. There is no reason to expect the Wilson–Butcher theory to give perfect agreement, since there are additional terms which are higher powers in $O(V_G^3)$. Similarly, most calculations of the Wilson–Butcher formula use the RPA form for $\varepsilon(G, \omega)$, whereas, the Singwi–Sjölander form would be an improvement. There is also a slight uncertainly as to the precise value of the pseudopotential V_{110}. With these qualifications in mind, the agreement between theory and experiment in Fig. 9.4 is satisfactory.

There is one difference between the original formula of Wilson–Butcher and the result (9.64) derived from the force–force correlation function. The original formula had the

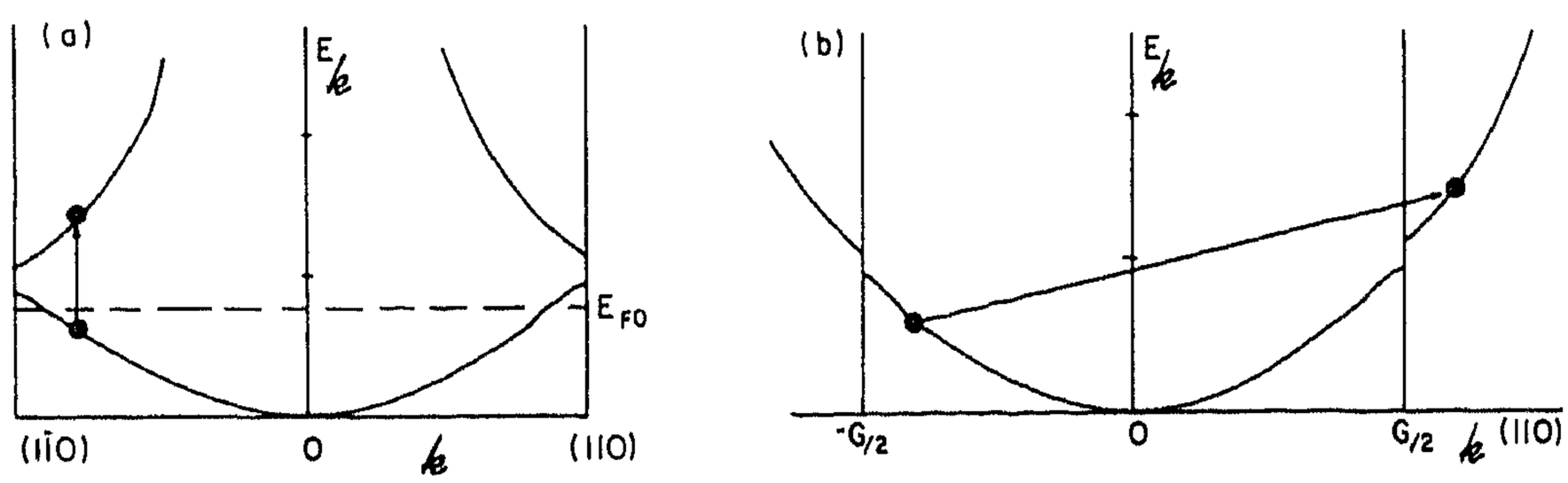


FIGURE 9.3 Interband transitions in alkali metals as shown in (a) reduced zone and (b) extended zone representations.

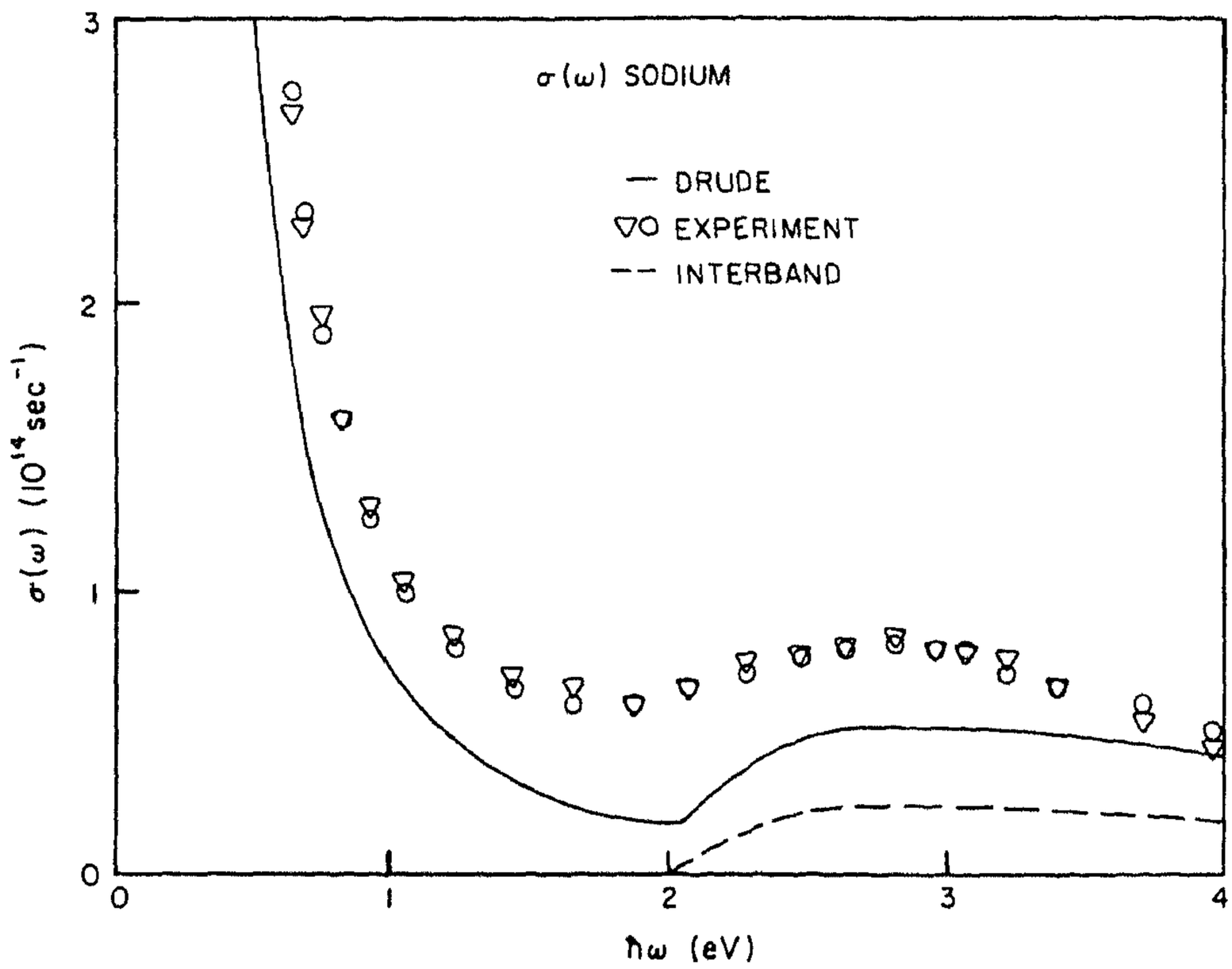


FIGURE 9.4 The optical conductivity for metallic sodium. The points are the experiments of Smith, while the lines are the theoretical contribution from Drude and interband transitions. *Source*: Smith (1969).

electron ion potential divided by the static dielectric function $\varepsilon(G, 0)$. The present derivation by the force–force correlation function is exact to order V_G^2, so its factor of $\varepsilon(G, \omega)$ is correct, and the old formula is not. The difference is more philosophical than numerical for sodium, since the frequency dependence of $\varepsilon(G, \omega)$ is slight over the range of frequencies shown in Fig. 9.4. Actually these frequencies, are not much lower than the plasma frequency at $\hbar\omega_p \approx 5$ eV, but this fact is irrelevant, since at these large values of wave vector G the dielectric function becomes frequency dependent only at much larger values of ω.

The alkali metals are good examples of where the Wilson–Butcher formula works well. Here the lowest interband transition is in the visible region of the spectrum and is isolated in frequency from any other feature except the Drude contribution, which is always present. Ashcroft and Sturm (1971) have shown that another formula is needed for polyvalent metals such as aluminum, since there are transitions between parallel bands which cause sharp structure at lower photon frequencies.

9.1.5. Phonons

Equation (9.25) lists three contributions to $\Phi(\mathbf{q})$. The first two terms are Bragg scattering and impurity scattering, which have been discussed. The third term is treated here. It is due to the electron scattering by the phonons.

Equation (9.35) is evaluated, with $\Phi(\mathbf{q})$ being the electron–phonon interaction. Only terms are retained that are of $O(M_\lambda^2)$. This point is rather tricky, since a subset of higher-order contributions are also retained. It is assumed that the phonon modes are the actual modes in the solid. The calculation of these modes includes electron–electron and electron–phonon interactions, which means they include M_λ. Any time the actual phonon modes are used, then one is including a subset of higher-order diagrams.

In Eq. (9.35) the phonon part is

$$-\langle T_\tau \Phi(\mathbf{q},\tau)\Phi(-\mathbf{q},0)\rangle = \frac{1}{\nu}\sum_\lambda M_\lambda^2 \mathscr{D}(\mathbf{q},\tau) \tag{9.65}$$

Equation (9.35) is now evaluated. Only keep the term that depends upon frequency $i\omega$:

$$\tau(i\omega) = -\frac{e^2}{\nu m^2 (i\omega)^2}\sum_{\mathbf{q}\lambda}\frac{q_\mu^2 M_\lambda(\mathbf{q})^2}{v_q}\frac{1}{\beta}\sum_{iq}\mathscr{D}_\lambda(\mathbf{q}, iq)\left[\frac{1}{\varepsilon(\mathbf{q}, i\omega - iq)} - 1\right]$$

The summation over Matsubara frequencies is evaluated using the techniques of Sec. 3.5. The phonon Green's function is expressed in the Lehmann representation using its spectral function $B(\mathbf{q},\omega)$. The inverse dielectric function is expressed similarly, with its spectral function being $\mathrm{Im}(1/\varepsilon)$

$$\tau(i\omega) = -\frac{1}{4\pi\nu m^2(i\omega)^2}\sum_{\mathbf{q}\lambda} q_\mu^2 q^2 M_\lambda(\mathbf{q})^2 \int\frac{d\omega'}{\pi}B(\mathbf{q},\omega')\int\frac{d\varepsilon'}{2\pi}\mathrm{Im}\left[\frac{1}{\varepsilon(\mathbf{q},\varepsilon')}\right]\frac{n_B(\omega') - n_B(\varepsilon')}{i\omega + \omega' - \varepsilon'} \tag{9.66}$$

The retarded function is obtained by taking $i\omega \to \omega + i\delta$. Then take the imaginary part. These steps cause the denominator $(i\omega + \omega' - \varepsilon')$ to be replaced by the numerator $\pi\delta(\omega + \omega' - \varepsilon')$. The delta function is used to eliminate the integral over $d\varepsilon'$.

This expression is usually evaluated at relatively low frequencies. When ω is much less than the plasma frequency of the metal, use the low-frequency limit for $\varepsilon_2(\mathbf{q},\omega) = 2\omega m^2 e^2 q^{-3}\Theta(2k_F - q)$. The various terms involving the phonon wave vector, matrix element, and spectral function are just the definition of the transport form of $\alpha_{tr}^2 F(u)$. The remaining factors combine to give

$$\mathrm{Re}[\sigma(\omega)] = -\frac{\mathrm{Im}[\pi(\omega)]}{\omega} \tag{9.67}$$

$$= \frac{2\pi e^2 n_0}{m\omega^3}\int_0^\infty du\,\alpha_{tr}^2 F(u)\{(\omega - u)[n_B(u) - n_B(u-\omega)]$$

$$+ (\omega + u)[n_B(u) - n_B(u+\omega)]\} \tag{9.68}$$

At zero temperature, the boson functions n_B arrange to limit the range of integration to $0 < u < \omega$. This limit gives the simple formula first derived by Allen (1971):

$$\mathrm{Re}[\sigma(\omega)] = \frac{2\pi e^2 n_0}{m\omega^3}\int_0^\omega du\,\alpha_{tr}^2 F(u)(\omega - u) \tag{9.69}$$

The Allen formula is approximate, since it is based upon the force–force correlation function. A more accurate expression would be based upon solving the Kubo relation for the conductivity as in Sec. 8.4. This derivation gives the result

$$\mathrm{Re}[\sigma(\omega)] = -\frac{e^2 n_0}{m\omega}\int_{-\infty}^\infty d\varepsilon[n_F(\varepsilon + \omega) - n_F(\varepsilon)] \tag{9.70}$$

$$\times\,\mathrm{Re}[\gamma(\varepsilon - i\delta, \varepsilon + \omega + i\delta)I(\varepsilon, \varepsilon + \omega)] \tag{9.71}$$

$$I(\varepsilon,\varepsilon') = \frac{2\pi i}{\Omega(\varepsilon) - \Omega(\varepsilon') + i[\Gamma(\varepsilon) + \Gamma(\varepsilon')]} \tag{9.72}$$

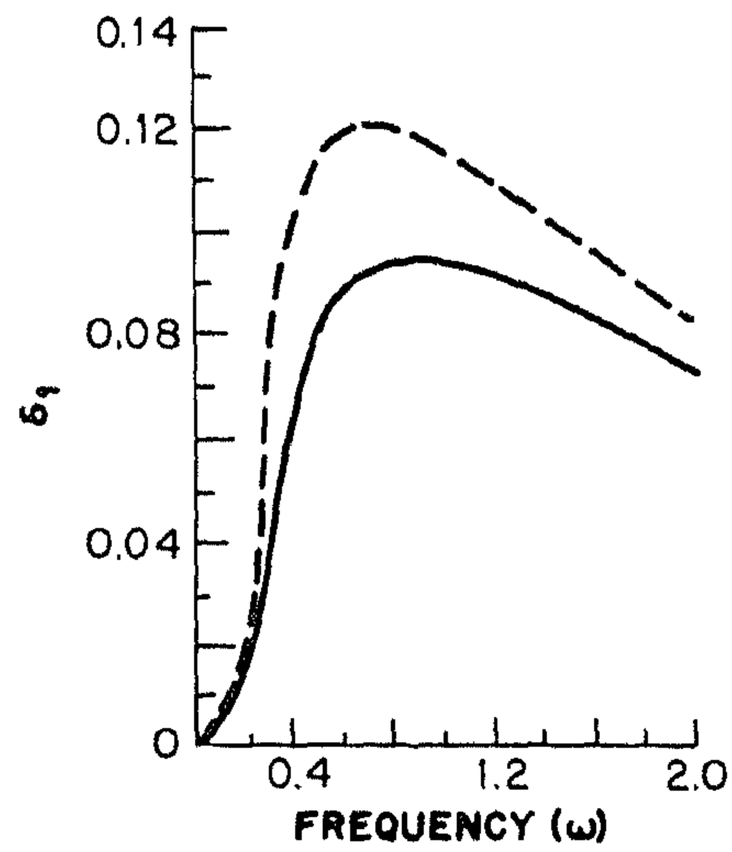


FIGURE 9.5 Phonon-induced absorption in calcium as a function of frequency. The vertical axis is $\mathrm{Re}[\sigma(\omega)m\omega/e^2n_0]$. The horizontal axis is ω/ω_D, where the Debye frequency is ω_D. The solid line is the solution to the ac transport equation, while the dashed line is using the Allen formula from the force–force correlation function. The latter method is approximate. *Source*: Wu and Mahan (1984).

This result can be derived directly from (8.204) and (8.213). The vertex function $\gamma(\varepsilon - i\delta, \varepsilon + \omega + i\delta)$ obeys the integral equation (8.228). A comparison of these two expressions is given in Fig. 9.5 for metallic calcium with $E \| c$. What is plotted is $\mathrm{Re}(\sigma/\sigma_0)$ vs. ω/ω_D, where $\sigma_0 = n_0e^2/m\omega$. The solid line is the exact result from (9.70) while the dashed line is the Allen formula in (9.69). The Allen formula has an error of about 10–15%. This makes it a good approximation, since (9.69) is very much easier to evaluate than (9.70). The latter requires solving numerically an integral equation for the vertex function.

9.2. WANNIER EXCITONS

9.2.1. The Model

Exciton states play an extremely important role in the understanding of interband transitions in semiconductors. The word *exciton* is used here to signify the modification of the absorption rate of photons due to the Coulomb interaction between the electron and the valence band hole. The easiest case to understand is when the lowest interband transition is direct: which means that the conduction band minimum and valence band maximum are at the same point in **k** space. This configuration is shown schematically in Fig. 9.6(a). The valence band states are all filled, and the conduction band states are all empty. The vertical arrow shows a possible interband transition which can occur when a photon is absorbed in the solid. The arrow starts in one of the valence states, since these are occupied by electrons, and goes to an unoccupied state in the conduction band. The valence state is shown as nondegenerate (except for spin) at $\mathbf{k} = 0$, although that is seldom the case; usually the band has an orbital degeneracy and is anisotropic.

The point of view in Fig. 9.6(a) is a single-particle picture of the transition process. According to this picture, the transition rate for the absorption of photons is given by the golden rule:

$$A(\omega) = \frac{2\pi}{\hbar\nu}\sum_{\mathbf{k}\mathbf{k}'} |\langle c, \mathbf{k}'|\hat{\varepsilon}\cdot\mathbf{p}|v, \mathbf{k}\rangle|^2\delta[\varepsilon_v(\mathbf{k}) + \hbar\omega - \varepsilon_c(\mathbf{k}')] \tag{9.73}$$

$$\varepsilon_v(\mathbf{k}) = -\frac{k^2}{2m_v}, \qquad \varepsilon_c(\mathbf{k}) = E_g + \frac{k^2}{2m_c} \tag{9.74}$$

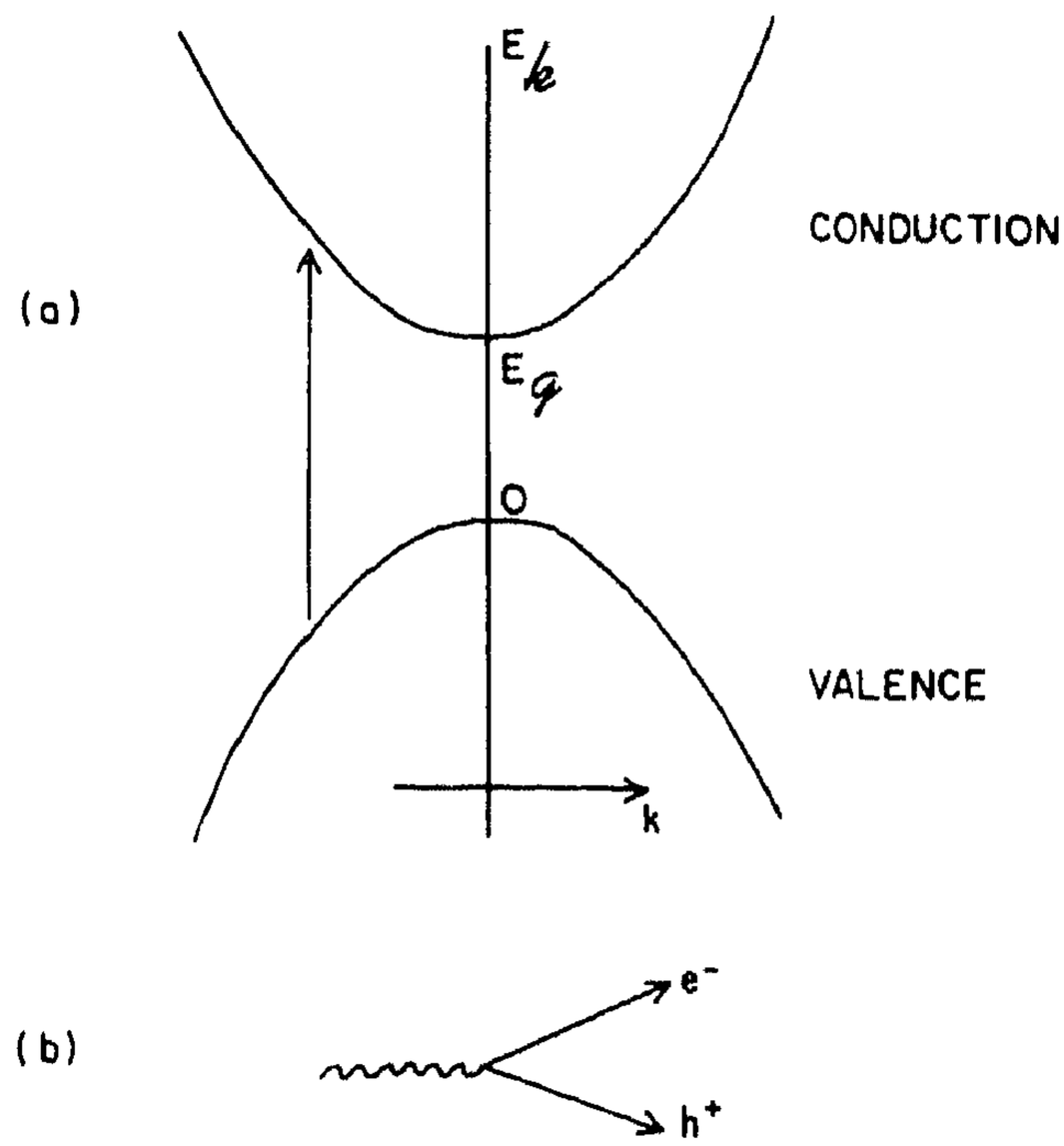


FIGURE 9.6 Optical transition in a semiconductor between the occupied valence band and empty conduction band for a direct transition. (a) Conventional band picture; (b) Wannier picture where the photon makes an electron-hole.

The energy zero is chosen to be the top of the valence band, so the bottom of the conduction band starts at the energy gap E_g. The matrix element $\langle c, \mathbf{k}'|\hat{\varepsilon} \cdot \mathbf{p}|v, \mathbf{k}\rangle$ is between the one-electron initial and final states. The wave functions are taken to be Bloch functions; which are the product of a cell-periodic part $u_{n\mathbf{k}}(\mathbf{r})$ and an envelope function $\exp(i\mathbf{k} \cdot \mathbf{r})$. It is a reasonable approximation, for our simple level of discussion, to take the cell-periodic parts as independent of wave vector $u_{c\mathbf{k}}(\mathbf{r}) = u_c(\mathbf{r})$, $u_{v\mathbf{k}}(\mathbf{r}) = u_v(\mathbf{r})$. It is also assumed that the valence band has p symmetry and the conduction band has s symmetry, which is true in many semiconductors but not all. Then $u_v(\mathbf{r})$ is a periodic orbital with angular momentum $l = 1$, while $u_c(\mathbf{r})$ is a periodic orbital with $l = 0$. With these approximations the optical matrix element is a constant except for wave vector conservation:

$$|v, \mathbf{k}\rangle = u_v(\mathbf{r}) \frac{e^{i\mathbf{k} \cdot \mathbf{r}}}{\sqrt{\mathcal{V}}} \tag{9.75}$$

$$|c, \mathbf{k}'\rangle = u_c(\mathbf{r}) \frac{e^{i\mathbf{k}' \cdot \mathbf{r}}}{\sqrt{\mathcal{V}}} \tag{9.76}$$

$$\langle c, \mathbf{k}'|\hat{\varepsilon} \cdot \mathbf{p}|v, \mathbf{k}\rangle = \langle c|\hat{\varepsilon} \cdot \mathbf{p}|v\rangle \delta_{\mathbf{k}\mathbf{k}'} \equiv \hat{\varepsilon} \cdot \mathbf{p}_{cv} \delta_{\mathbf{k}\mathbf{k}'} \tag{9.77}$$

$$\langle c|\hat{\varepsilon} \cdot \mathbf{p}|v\rangle = \frac{1}{\mathcal{V}_0} \int_{\text{cell}} d^3r u_c^* \hat{\varepsilon} \cdot \mathbf{p} u_v \tag{9.78}$$

where $\mathcal{V}_0$ is the volume of a unit cell. The matrix element $\langle c|\mathbf{p}|v\rangle$ should also contain the factor $\exp(i\mathbf{q} \cdot \mathbf{r})$, where $\mathbf{q}$ is the wave vector of the photon. This wave vector is small enough to be neglected for photons in the optical frequencies. Now that the matrix element is evaluated, it is possible to do the integrals over the electron wave vector:

$$\begin{aligned} A(\omega) &= \frac{2\pi}{\hbar} |\hat{\varepsilon} \cdot \mathbf{p}_{cv}|^2 \int \frac{d^3k}{(2\pi)^3} \delta\left(\hbar\omega - E_g - \frac{k^2}{2\mu}\right) \\ &= |\hat{\varepsilon} \cdot \mathbf{p}_{cv}|^2 \frac{(2\mu)^{3/2}}{2\pi} \sqrt{\hbar\omega - E_g}\, \Theta(\hbar\omega - E_g) \end{aligned} \tag{9.79}$$

$$\frac{1}{\mu} = \frac{1}{m_c} + \frac{1}{m_v} \tag{9.80}$$

Equation (9.79) predicts that the absorption rate begins at the energy gap of the semiconductor and rises as the square root of the factor $(\hbar\omega - E)_g$. It is a very definite prediction, which is not observed. In fact, the one-particle theory is totally inadequate and does not come close to describing the absorption spectra $A(\omega)$ observed in experiments.

Wannier (1937) observed that the interband transition in semiconductors was really a two-particle process, which is indicated in Fig. 9.6(b). Since the valence band states are all occupied, removing an electron from this band creates an excitation called a hole. This hole, which is the absence of an electron in an otherwise filled band, can be treated as a particle with an effective mass m_v and a positive charge. The energy of the hole is $\varepsilon_h(k) = k^2/2m_v$. In the two-particle picture, the photon of energy creates two excitations in the semiconductor: the electron in the conduction band of wave vector $\mathbf{k}$ and energy $E_g + k^2/2m_c$ and the hole in the valence band of wave vector $-\mathbf{k}$ and energy $k^2/2m_v$. Energy conservation is

$$\hbar\omega = E_g + \frac{k^2}{2m_c} + \frac{k^2}{2m_v} = E_g + \frac{k^2}{2\mu} \tag{9.81}$$

which is exactly the same as the argument of the delta function for energy conservation in (9.79). So far the two-particle picture leads to the same absorption rate as the one-particle prediction in (9.79). Both the matrix and the energy conservation are identical, so the prediction is exactly the same.

The point made by Wannier is that there is now additional physics which can occur in the two-particle picture. The electron and hole are particles with charges of opposite signs, so that there is a Coulomb attraction $-e^2/(\varepsilon_0 r)$ between them, where ε_0 is the static dielectric function. This dielectric function is assumed to be a constant which is independent of frequency. This approximation is poor, since most semiconductors are polar, and the dielectric function has significant dispersion at frequencies near the optical phonon frequencies; see Sec. 6.3.1. The frequency-dependent screening is an interesting problem which is ignored by treating ε_0 as a constant. The attractive Coulomb interaction between the electron and hole can cause hydrogenic bound states between them.

The optical absorption rate for this process was calculated by Elliott (1957). The final state of the system is described by a two-particle Schrödinger equation:

$$\Psi(\mathbf{r}_e, \mathbf{r}_h) = u_c(\mathbf{r}_e)u_v(\mathbf{r}_h)\Phi(\mathbf{r}_e, \mathbf{r}_h) \tag{9.82}$$

$$0 = \left[-\frac{\hbar^2\nabla_e^2}{2m_c} - \frac{\hbar^2\nabla_h^2}{2m_v} - \frac{e^2}{\varepsilon_0|\mathbf{r}_e - \mathbf{r}_h|} - E\right]\Phi(\mathbf{r}_e, \mathbf{r}_h) \tag{9.83}$$

$\Phi(\mathbf{r}_e, \mathbf{r}_h)$ can be factored into relative $\mathbf{r} = \mathbf{r}_e - \mathbf{r}_h$ and center of mass coordinates $M = m_c + m_v$ in the standard fashion:

$$\mathbf{R} = \frac{m_c\mathbf{r}_e + m_v\mathbf{r}_h}{M} \tag{9.84}$$

$$\Phi(\mathbf{r}_e, \mathbf{r}_h) = \frac{e^{i\mathbf{P}\cdot\mathbf{R}}}{\sqrt{\mathcal{V}}}\phi(\mathbf{r}) \tag{9.85}$$

$$\left(-\frac{\hbar^2}{2\mu}\nabla^2 - \frac{e^2}{\varepsilon_0 r} - \varepsilon_r\right)\phi(\mathbf{r}) = 0 \tag{9.86}$$

$$E = E_g + \varepsilon_r + \frac{P^2}{2M} \tag{9.87}$$

The center of mass motion is plane-wave-like, with a wave vector $\mathbf{P}$ which in optical experiments is equal to the photon wave vector. This wave vector is small. It is neglected in the center of mass motion and set $\mathbf{P} = 0$. The relative motion of the electron and hole is usually more interesting. For relative energy ε_r less than zero, the two particles form bound hydrogenic states with energies $\varepsilon_r = \varepsilon_n = -E_R/n^2$. For relative energy ε_r greater than zero they form scattering states $\phi_{\mathbf{k}}(\mathbf{r})$. Elliott showed that the optical transition rate depends on the relative wave function at $\mathbf{r} = 0$, $\phi_{\mathbf{k}}(0)$. Instead of (9.79), the transition rate is determined by

$$A(\omega) = \frac{2\pi}{\hbar} |\varepsilon \cdot \mathbf{p}_{cv}|^2 \sum_j |\phi_j(0)|^2 \delta(\hbar\omega - E_g - \varepsilon_j) \tag{9.88}$$

The summation j is over the bound and continuum states of the relative motion of the electron–hole pair. The dependence on the relative wave function, evaluated at $\mathbf{r} = 0$, can be understood by a physical argument. First, consider the corresponding emission process, whereby the electron and hole recombine to emit a photon with energy $\hbar\omega$. The emission rate should depend on the relative wave function at $\mathbf{r} = 0$, since that is the probability that the electron and hole find themselves at the same spot in the solid, which is necessary for the recombination. The dependence of the emission rate on $|\phi_j(0)|^2$ is reasonable from physical intuition. But the matrix elements for absorption and emission are identical, so that absorption should have the same dependence. The absorption depends on the probability of making the electron and hole at the same point in the solid.

The former result, for the noninteracting solid, is recovered from (9.88) by setting $|\phi_j(0)|^2 = 1/\nu$ and $\varepsilon_j = k^2/2\mu$. The results, for the interacting case are quite different. The relative motions of the electron and hole are in s-wave hydrogenic states, either bound or unbound, because of the angular momentum selection rule. The one-unit change in l, in the photon absorption, is taken by the change of band symmetry, and the relative motion is not permitted any additional angular momentum. For s states, the bound states have an amplitude given by the principal quantum number n and the Bohr radius a_0:

$$\phi_n(0) = \frac{1}{\sqrt{\pi a_0^3 n^3}} \tag{9.89}$$

For continuum states, with energy $\varepsilon_k = k^2/2\mu$, the relative wave function at the origin is $|\phi_{k,l=0}(0) = 2\pi\eta/\nu[1 - \exp(-2\pi\eta)]$, where $1/\eta = ka_0$. Then (9.88) predicts that the absorption is a constant in frequency at the energy gap E_g, and does not rise with a square root dependence on $(\hbar\omega - E_g)$. The actual shape is shown in Fig. 9.7. A few sharp, distinct exciton lines are observed at low frequency which correspond to 1s, 2s, etc. In actual experiments, these absorption bands are hard to see because they are too strong. Only in the thinnest crystals can enough light be transmitted through the experimental sample to measure the absorption rate at the bound states of the excitons. Usually the absorption is so strong that all the light is attenuated before transversing the sample. At higher frequencies, the ns states

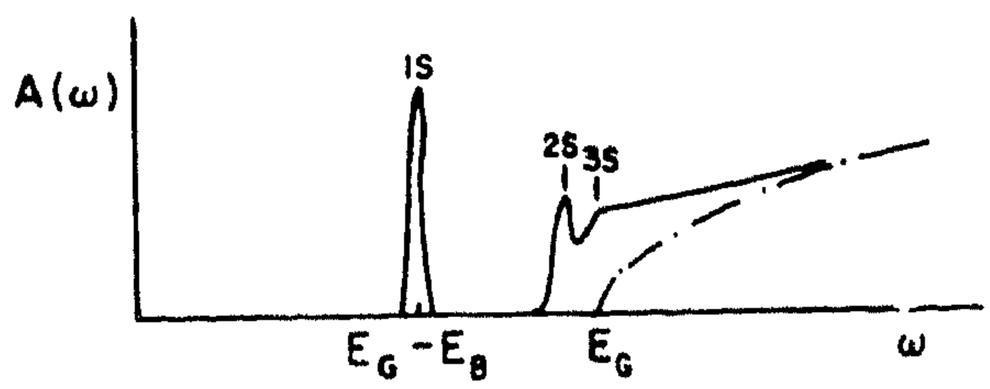


FIGURE 9.7 Typical absorption spectra of a direct gap semiconductor. Distinct exciton lines are labeled 1s, 2s, etc. The solid line is the theory with final state interactions, and the dashed line is the theory without them.

are closer in frequency and are broadened, so that their envelope becomes a continuous distribution which merges with the continuum absorption which starts at $\hbar\omega = E_g$. This experiment shows no anomaly at $\hbar\omega = E_g$, and the experimental value of E_g is obtained only by extrapolating the Rydberg series for the positions of the ns exciton lines:

$$\omega_n = E_g + \varepsilon_n = E_g - \frac{E_R}{n^2} \tag{9.90}$$

In most semiconductors, the Rydberg unit E_R is of the order of several millivolts because of the high values of the dielectric constant $\varepsilon_0 \approx 10$ and the small values of the effective mass $\mu \approx 0.1$.

This discussion is only a brief review of exciton theory in semiconductors. The interested reader is encouraged to seek additional information in the books by Knox (1963) and Reynolds and Collins (1981). The theory has been verified by many detailed experiments, such as exciton properties in static electric or magnetic fields, under stress, etc. Exciton effects are important for determining the absorption rate of photons within a frequency interval of E_R near the energy gap E_g.

9.2.2. Solution by Green's Functions

The Elliott formula (9.88) is derived using Green's functions. The starting point is the Hamiltonian (9.86) which contains the properties of electrons, holes, and their mutual interaction. The Kubo formula is evaluated for the optical conductivity, which is proportional to $A(\omega)$. The Elliott forumula is obtained by a summation of all the vertex diagrams of the current–current correlation function. The Green's function analysis is straightforward and can be solved exactly, because it is still only a two-body problem. Indeed, this section could have been included in Chapter 4 on exactly solvable models. The motivation for this analysis is to stress the importance of final state interactions. The latter is a fancy name for the interactions between the particles in the final state of the optical transition. In this case they are the Coulomb interaction between the electron and hole.

The starting point for the Green's function calculation is the Kubo formula for the current–current correlation function:

$$\pi(i\omega) = -\frac{1}{\nu}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau j_\mu(\tau) j_\mu(0)\rangle \tag{9.91}$$

$$j_\mu = \sum_{\mathbf{k}\sigma} w_{\mathbf{k}}(C_{\mathbf{k}\sigma} d_{-\mathbf{k}-\sigma} + d^\dagger_{-\mathbf{k}-\sigma} C^\dagger_{\mathbf{k}\sigma}) \tag{9.92}$$

$$w_{\mathbf{k}} = \int d^3r \Psi^*_c(\mathbf{k}, \mathbf{r}) p_\mu \Psi_v(\mathbf{k}, \mathbf{r}) = \langle c, \mathbf{k}|p_\mu|v, \mathbf{k}\rangle \tag{9.93}$$

The τ development of the operators is governed by a Hamiltonian which contains the Coulomb interaction $v_q = -4\pi e^2/\varepsilon_0 q^2$ between the electron $(C_{\mathbf{k}\sigma}, C^\dagger_{\mathbf{k}\sigma})$ and the hole $(d_{\mathbf{k}\sigma}, d^\dagger_{\mathbf{k}\sigma})$

$$H = H_0 + V \tag{9.94}$$

$$H_0 = \sum_{\mathbf{k}} \xi_v(\mathbf{k}) d^\dagger_{\mathbf{k}} d_{\mathbf{k}} + \sum_{\mathbf{k}} \xi_c(\mathbf{k}) C^\dagger_{\mathbf{k}} C_{\mathbf{k}} \tag{9.95}$$

$$V = \frac{1}{\nu}\sum_{\mathbf{q}} v_q \left(\sum_{\mathbf{k}} C^\dagger_{\mathbf{k}+\mathbf{q}} C_{\mathbf{k}}\right)\left(\sum_{\mathbf{p}} d^\dagger_{\mathbf{p}} d_{\mathbf{p}}\right) \tag{9.96}$$

Spin does not play a role in this analysis, so the spin index is no longer written. The procedure is to solve the Kubo formula (9.92) with this Hamiltonian. The optical matrix element $w_{\mathbf{k}}$ is taken here to be a function of $\mathbf{k}$, although in the theory for Wannier excitons it is a constant. An important assumption, when evaluating this correlation function, is that the densities of electrons n_e and holes n_h are both negligibly small. An equivalent assumption is that the chemical potential μ in the semiconductor is in the forbidden energy gap, somewhere between energy zero and E_g. The chemical potential μ_h for the holes is the negative of that for the electrons $\mu_h = -\mu_e$, and the energies in the preceding Hamiltonian are

$$\xi_e(\mathbf{k}) = \frac{k^2}{2m_c} + E_g - \mu \tag{9.97}$$

$$\xi_h(\mathbf{k}) = \frac{k^2}{2m_v} - \mu_h = \frac{k^2}{2m_v} + \mu \tag{9.98}$$

The definitions are finished, and it is time to evaluate the correlation function.

The electron–hole interaction V is treated as the perturbation for the Hamiltonian. The first term is that for no interaction, which is called $\pi^{(0)}(i\omega)$:

$$\pi^{(0)}(i\omega) = -\frac{1}{\nu}\sum_{\mathbf{k}} w_{\mathbf{k}}^2 \int_0^\beta d\tau e^{i\omega_n\tau} \langle T_\tau C_{\mathbf{k}}(\tau) d_{-\mathbf{k}}(\tau) d_{-\mathbf{k}}^\dagger(0) C_{\mathbf{k}}^\dagger(0)\rangle \tag{9.99}$$

$$= -\frac{1}{\nu}\sum_{\mathbf{k}} w_{\mathbf{k}}^2 \int_0^\beta d\tau e^{i\omega_n\tau} \mathscr{G}_c(\mathbf{k}, \tau)\mathscr{G}_h(-\mathbf{k}, \tau) \tag{9.100}$$

$$= -\frac{1}{\nu}\sum_{\mathbf{k}} w_{\mathbf{k}}^2 \frac{1}{\beta}\sum_{ik} \mathscr{G}_c(\mathbf{k}, ik)\mathscr{G}_h(-\mathbf{k}, i\omega - ik) \tag{9.101}$$

$$= \frac{1}{\nu}\sum_{\mathbf{k}} w_k^2 \frac{1 - n_F(\xi_e) - n_F(\xi_h)}{i\omega - \xi_e(k) - \xi_h(k)} \tag{9.102}$$

The last step involves a standard frequency summation, as given in Sec. 3.5. Both Green's functions $\mathscr{G}_c = 1/(ik - \xi_e)$ and $\mathscr{G}_h = 1/(ik - \xi_h)$ are for noninteracting particles, but the superscripts (0) are omitted since all the Green's functions in this section are noninteracting. The occupation factors n_F are zero. The absorption $A(\omega)$ is the spectral function of this correlation function. Take the analytical continuation $i\omega \to \omega + i\delta$ and define

$$A(\omega) = -2\,\mathrm{Im}[\pi(\omega)] \tag{9.103}$$

$$A^{(0)}(\omega) = -2\,\mathrm{Im}[\pi^{(0)}(\omega)] = 2\pi \int \frac{d^3k}{(2\pi)^3} w_{\mathbf{k}}^2 \delta(\omega - \xi_v(\mathbf{k}) - \xi_e(\mathbf{k}))$$

$$= 2\pi \int \frac{d^3k}{(2\pi)^3} w_{\mathbf{k}}^2 \delta\left(\omega - E_g - \frac{k^2}{2\mu}\right) \tag{9.104}$$

The result for $A^{(0)}(\omega)$ is recognized as the one-particle theory of (9.79). The remaining terms in the expansion of the S matrix are now examined. Fortunately, most of them are zero. All self-energy terms, for either the electron or the hole, are zero. All self-energy terms for the electron or hole contain closed loops and vanish since they are proportional to n_F. The only nonzero contributions are vertex diagrams in the form of ladder diagrams.

The method of summing the ladder diagrams follows Mahan (1967) and was used for impurity scattering in Sec. 4.1.5. They are shown as the diagrams in Fig. 9.8. The first vertex correction $A^{(1)}$ contains a single ladder diagram, where the dashed interaction represents the

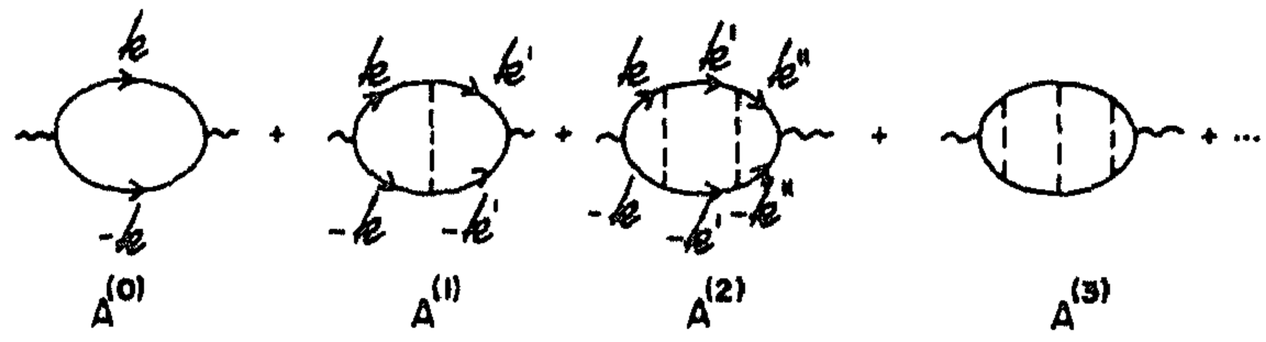


FIGURE 9.8

Coulomb interaction between the electron and the hole. The two particles are represented by the solid line segments. The wiggly line at the vertex represents the photon. The lines for both the electron and hole are shown coming from the photon vertex and are traveling parallel in time. The diagram with n ladders is called $A^{(n)}$. They can be constructed according to the rules for diagram construction. The first ladder diagrams gives

$$\pi^{(1)}(i\omega) = \frac{1}{\nu^2}\sum_{\mathbf{k}\mathbf{k}'} w_{\mathbf{k}} v_{\mathbf{k}-\mathbf{k}'} w_{\mathbf{k}'} \frac{1}{\beta}\sum_{ik} \mathcal{G}_e(\mathbf{k}, ik)\mathcal{G}_h(-\mathbf{k}, i\omega - i\xi)$$
$$\times \frac{1}{\beta}\sum_{ik'} \mathcal{G}_e(\mathbf{k}', ik')\mathcal{G}_h(-\mathbf{k}', i\omega - ik') \tag{9.105}$$

The two summations over Matsubara frequency are identical in form:

$$\frac{1}{\beta}\sum_{ik} \mathcal{G}_e(\mathbf{k}, ik)\mathcal{G}_h(-\mathbf{k}, i\omega - ik) = \frac{1 - n_F(\xi_e) - n_F(\xi_h)}{i\omega - \xi_e(k) - \xi_h(\mathbf{k})}$$
$$= \frac{1}{i\omega - E_g - k^2/2\mu}$$
$$\pi^{(1)}(i\omega) = \frac{1}{\nu^2}\sum_{\mathbf{k}\mathbf{k}'} w_{\mathbf{k}} v_{\mathbf{k}-\mathbf{k}'} w_{\mathbf{k}'} \frac{1}{i\omega - E_g - k^2/2\mu}\frac{1}{i\omega - E_g - k'^2/2\mu} \tag{9.106}$$

Additional ladder diagrams in $A^{(n)}$ just produce additional factors of the Green's function pair $\mathcal{G}_c\mathcal{G}_h$, which are summed over Matsubara frequency to produce another energy denominator $(i\omega - E_g - k_j^2/2\mu)$. Each ladder also has another factor of the Coulomb potential $v_{\mathbf{k}_i-\mathbf{k}_j}/\nu$. For this problem it is easy to write the term with n ladder diagrams:

$$\pi^{(n)} = \frac{1}{\nu^{n+1}}\sum_{\mathbf{k}\mathbf{k}_1\cdots\mathbf{k}_n} \frac{w_{\mathbf{k}} v_{\mathbf{k}-\mathbf{k}_1}\cdots v_{\mathbf{k}_{n-1}-\mathbf{k}_n} w_{\mathbf{k}_n}}{i\omega - E_g - k^2/2\mu}\Pi_j \frac{1}{i\omega - E_g - k_j^2/2\mu}$$

The simple form of these terms permits them to be summed exactly to obtain the expression for the correlation function. The other vertex diagrams, which are not ladder diagrams but have a crossing of the interaction lines, are all proportional to the numbers of electrons or holes in the semiconductor. They can be neglected, and the only nonzero vertex corrections are the ladder diagrams.

The summation of the ladder diagrams can be expressed as a vertex correction to the correlation function, which is written as

$$\pi_L(i\omega) = \sum_n \pi^{(n)} = \frac{1}{\nu}\sum_{\mathbf{k}} \frac{w_{\mathbf{k}}\Gamma(\mathbf{k}, i\omega)}{i\omega - E_g - k^2/2\mu} \tag{9.107}$$

$$\Gamma(\mathbf{k}, i\omega) = w_{\mathbf{k}} + \frac{1}{\nu}\sum_{\mathbf{k}_1} \frac{v_{\mathbf{k}-\mathbf{k}_1}\Gamma(\mathbf{k}_1, i\omega)}{i\omega - E_g - k_1^2/2\mu} \tag{9.108}$$

An exact summation of the ladder diagrams is obtained by solving for this vertex function. The integral equation (9.108) for the vertex function is an equation of the relative motion of the two particles.

The first step in the solution of the integral equation (9.108) is to define the auxiliary function

$$P(\mathbf{r}, i\omega) = \frac{1}{\nu}\sum_{\mathbf{k}} \frac{e^{i\mathbf{k}\cdot\mathbf{r}}\Gamma(\mathbf{k}, i\omega)}{i\omega - E_g - k^2/2\mu} \tag{9.109}$$

For the Wannier exciton system, the matrix element $w_{\mathbf{k}}$ is independent of $\mathbf{k}$ and is a constant w_0. In this case, the correlation function in (9.107) is the value of this auxiliary function at the origin:

$$\pi_L(i\omega) = w_0 P(\mathbf{r} = 0, i\omega) \tag{9.110}$$

The integral equation (9.108) for the vertex function becomes

$$\Gamma(\mathbf{k}, i\omega) = w_0 - \int d^3r \frac{e^2}{\varepsilon_0 r} e^{-i\mathbf{k}\cdot\mathbf{r}} P(\mathbf{r}, i\omega) \tag{9.111}$$

The auxiliary function $P(\mathbf{r}, i\omega)$ is introduced so that one can write a differential equation for it, which is then solved to obtain the exact solution to the function. This differential equation is obtained by operating on both sides of (9.109) with the operator $[i\omega - E_g + \nabla^2/2\mu]$, which is chosen to eliminate the energy denominator on the right-hand side:

$$\left(i\omega - E_g + \frac{\nabla^2}{2\mu}\right)P(\mathbf{r}, i\omega) = \frac{1}{\nu}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\Gamma(\mathbf{k}, i\omega)\frac{i\omega - E_g - k^2/2\mu}{i\omega - E_g - k^2/2\mu} \tag{9.112}$$

$$= \frac{1}{\nu}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\Gamma(\mathbf{k}, i\omega) \tag{9.113}$$

$$= \frac{1}{\nu}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}\left[w_0 - \int d^3r' \frac{e^2}{\varepsilon_0 r'} e^{-i\mathbf{k}\cdot\mathbf{r}'} P(\mathbf{r}', i\omega)\right]$$

$$= w_0\delta^3(\mathbf{r}) - \frac{e^2}{\varepsilon_0 r} P(\mathbf{r}, i\omega) \tag{9.114}$$

The integral equation (9.111) for the vertex function was used to obtain a differential equation for the function $P(\mathbf{r}, i\omega)$. The interaction term on the right is moved to the left-hand side of the equation, which derives the inhomogeneous differential equation:

$$(i\omega - E_g - H_{ex})P(\mathbf{r}, i\omega) = w_0\delta^3(\mathbf{r}) \tag{9.115}$$

$$H_{ex} = -\frac{\hbar^2}{2\mu}\nabla^2 - \frac{e^2}{\varepsilon_0 r} \tag{9.116}$$

This equation is solved in the following fashion. The factors $i\omega - E_g$ on the left are constants, and the important operator parts are the same as found in the exciton Hamiltonian (9.86) for their relative motion:

$$H_{ex}\phi_j(\mathbf{r}) = \varepsilon_j\phi_j(\mathbf{r}) \tag{9.117}$$

The Hamiltonian H_{ex} may be solved exactly in terms of hydrogenic eigenfunctions $\phi_j(\mathbf{r})$ and eigenvalues ε_j. In terms of these exact solutions, the solution to the differential equation (9.115) is written as

$$P(\mathbf{r}, i\omega) = w_0 \sum_j \frac{\phi_j(\mathbf{r})\phi_j^*(0)}{i\omega - E_g - \varepsilon_j} \tag{9.118}$$

which is verified by examining the original equation:

$$\begin{aligned}(i\omega - E_g - H_{ex})P(\mathbf{r}, i\omega) &= w_0 \sum_j \phi_j(\mathbf{r})\phi_j^*(0)\frac{i\omega - E_g - \varepsilon_j}{i\omega - E_g - \varepsilon_j} \\ &= w_0 \sum_j \phi_j(\mathbf{r})\phi_j^*(0) = w_0\delta^3(\mathbf{r})\end{aligned} \tag{9.119}$$

The last step is a summation over the complete set of hydrogenic states, which produces a delta function. This result agrees with (9.115), so that an exact solution is found for $P(\mathbf{r}, i\omega)$. The current–current correlated function $\pi_L(i\omega)$ is obtained from (9.110) as the value of $w_0P(\mathbf{r} = 0, i\omega)$:

$$\pi_L(i\omega) = w_0^2 \sum_j \frac{|\phi_j(0)|^2}{i\omega - E_g - \varepsilon_j} \tag{9.120}$$

The optical absorption function $A(\omega)$ is the spectral function of this operator:

$$A(\omega) = 2\pi w_0^2 \sum_j |\phi_j(0)|^2 \delta(\omega - E_g - \varepsilon_j)$$

This result is exactly the Elliott form in (9.88), with the transition rate depending on the square of the relative wave function evaluated at $\mathbf{r} = 0$. As discussed in Sec. 8.2.1, this dependence causes a dramatic and important change in the absorption spectra. The final state interactions are very important. They are obtained by summing the ladder diagrams of the correlation function. It is an exact solution to the current–current correlation function (9.92) for the model Hamiltonian in (9.96) and with the restriction that there is a negligible density of electrons and holes in the semiconductor.

9.2.3. Core-Level Spectra

Core-level spectra in insulators is another problem which is similar, both physically and mathematically, to the Wannier exciton. The initial electronic state in the optical transition is a core level of the atom, such as the 1s or 2s shell. The optical transition in the insulator takes the core electron to an unoccupied state in the conduction band. The empty core level can still be described as a hole, but now it is very localized in space and so has an infinite effective mass. The hole state has a constant energy $\xi_h = E_c + \mu$, where E_c is the core-level binding energy, as measured from the top of the valence band. The threshold for the optical transition, in the absence of any final state interactions, is the frequency $\omega_T = E_g + E_c$.

This problem can be solved exactly by the same techniques which were used for Wannier excitons. The advantage of the present model is that it is possible to explain the physics with greater clarity. From the point of view of the conduction electron, it has two possible potential functions. There is a potential for the initial state of the system, which is called $V_i(\mathbf{r})$. It is the potential acting on a conduction electron in the initial state, when there is no core hole. The

effective Hamiltonian for the conduction electron is called H_i, and its eigenfunctions are $\Psi_i(\mathbf{k}, \mathbf{r})$ with energy $\varepsilon_i(\mathbf{k})$:

$$H_i\Psi_i = \left[-\frac{\hbar^2}{2m}\nabla^2 + V_i(\mathbf{r})\right]\Psi_i(\mathbf{k}, \mathbf{r}) = \varepsilon_i(\mathbf{k})\Psi_i(\mathbf{k}, \mathbf{r}) \tag{9.121}$$

After the optical transition has occurred, the hole appears in the core state, which alters the potential of the conduction electron to $V_f(\mathbf{r})$. This final state potential is set equal to the initial state potential plus a term due to the core hole:

$$V_f = V_i + V_h \tag{9.122}$$

$$H_f\Psi_f = \left[-\frac{\hbar^2}{2m}\nabla^2 + V_f(\mathbf{r})\right]\Psi_f(\mathbf{k}, \mathbf{r}) = \varepsilon_f(\mathbf{k})\Psi_f(\mathbf{k}, \mathbf{r}) \tag{9.123}$$

The core hole potential $V_h(\mathbf{r})$ has the form of a Coulomb potential $-e^2/(\varepsilon_0 r)$ at long range, but there are atomic effects at short range. The final state Hamiltonian H_f has its own set of eigenfunctions $\Psi_f(\mathbf{k}, \mathbf{r})$ and eigenstates $\varepsilon_f(\mathbf{k})$ which are different from those of the initial state. These states may be bound or unbound, to the core hole, and $\mathbf{k}$ is a general index which can represent either of these possibilities. The conduction electron has two possible sets of eigenstates which could enter into the calculation for the absorption. Which one should be used?

The conventional approach to this problem is to expand the Hamiltonian and the current operator in the basis states of the initial Hamiltonian H_i. A set of creation and destruction operators $(C_{\mathbf{k}s}, C^\dagger_{\mathbf{k}s})$ are associated with the basis set Ψ_i. The effective Hamiltonian is

$$\begin{aligned} H &= \xi_h \sum_s d_s^\dagger d_s + \sum_{\mathbf{k}s} \xi_i(\mathbf{k}) C^\dagger_{\mathbf{k}s} C_{\mathbf{k}s} + \sum_{\mathbf{k}\mathbf{k}'ss'} V_h(\mathbf{k}, \mathbf{k}') C^\dagger_{\mathbf{k}s} C_{\mathbf{k}'s} d^\dagger_{s'} d_{s'} \\ \xi_i(\mathbf{k}) &= \varepsilon_i(\mathbf{k}) - \mu \end{aligned} \tag{9.124}$$

$$V_h(\mathbf{k}, \mathbf{k}') = \int d^3r \Psi_i^*(\mathbf{k}, \mathbf{r}) V_h(\mathbf{r}) \Psi_i(\mathbf{k}', \mathbf{r}) \tag{9.125}$$

This Hamiltonian has the same general form as (9.96). The last term in the Hamiltonian is the interaction between the conduction electron and the hole; the latter is represented by the operators $(d_s, d_s^\dagger)$. The current operator, for the Kubo formula (9.92), is also written in terms of these functions:

$$j_\mu = \sum_{\mathbf{k}s} w_i(\mathbf{k})(C^\dagger_{\mathbf{k}s} d_s^\dagger + d_s C_{\mathbf{k}s}) \tag{9.126}$$

$$w_i(\mathbf{k}) = \int d^3r \Psi_i(\mathbf{k}, \mathbf{r}) \hat{\varepsilon} \cdot \mathbf{p} \phi_c(\mathbf{r}) \tag{9.127}$$

where $\phi_c(\mathbf{r})$ is the core wave function of the hole.

The problem has the same form as the Wannier exciton which was solved previously and is treated similarly. The correlation function, from the Kubo formula, is evaluated as a set of

Max - Planck - Ir
für Mikrostruktur
Bibliothek

ladder diagrams. Their summation is achieved by a set of equations which are nearly identical to (9.107) and (9.108):

$$\pi(i\omega) = \frac{1}{\nu}\sum_{\mathbf{k}s} \frac{w_i(\mathbf{k})^*\Gamma_i(\mathbf{k}, i\omega)}{i\omega - \omega_T - \varepsilon_i(\mathbf{k})} \tag{9.128}$$

$$\Gamma_i(\mathbf{k}, i\omega) = w_i(\mathbf{k}) + \sum_{\mathbf{k}_1} \frac{V_h(\mathbf{k}, \mathbf{k}_1)\Gamma_i(\mathbf{k}_1, i\omega)}{i\omega - \omega_T - \varepsilon_i(\mathbf{k}_1)} \tag{9.129}$$

The vertex function $\Gamma_i(\mathbf{k}, i\omega)$ is solved by introducing the auxiliary function

$$P(\mathbf{r}, i\omega) = \sum_{\mathbf{k}} \frac{\Psi_i(\mathbf{k}, \mathbf{r})\Gamma_i(\mathbf{k}, i\omega)}{i\omega - \omega_T - \varepsilon_i(\mathbf{k})} \tag{9.130}$$

$$\pi(i\omega) = \frac{1}{\nu}\int d^3r\phi_c(\mathbf{r})\hat{\varepsilon}\cdot\mathbf{p}P(\mathbf{r}, i\omega) \tag{9.131}$$

The equation for the vertex function is

$$\Gamma_i(\mathbf{k}, i\omega) = \int d^3r\Psi_i^*(\mathbf{k}, \mathbf{r})[\hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}) + V_h(\mathbf{r})P(\mathbf{r}, i\omega)] \tag{9.132}$$

The correlation function $P(\mathbf{r}, i\omega)$ is solved by writing it as a differential equation, which is found by first operating by

$$\begin{aligned}(i\omega - \omega_T - H_i)P(\mathbf{r}, i\omega) &= \sum_{\mathbf{k}} \Psi_i(\mathbf{k}, \mathbf{r})\Gamma_i(\mathbf{k}, i\omega) \\ &= \sum_{\mathbf{k}} \Psi_i(\mathbf{k}, \mathbf{r})\int d^3r'\Psi_i^*(\mathbf{k}, \mathbf{r}')[\hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}') + V_h(\mathbf{r}')P(\mathbf{r}', i\omega)] \\ &= \int d^3r\delta(\mathbf{r} - \mathbf{r}')[\hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}') + V_h(\mathbf{r}')P(\mathbf{r}', i\omega)] \\ &= \hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}) + V_h(\mathbf{r}')P(\mathbf{r}, i\omega)\end{aligned} \tag{9.133, 9.134}$$

where the completeness relation is

$$\sum_{\mathbf{k}} \Psi_i(\mathbf{k}, \mathbf{r})\Psi_i^*(\mathbf{k}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \tag{9.135}$$

This relation produces the delta function which eliminates the integration over $\mathbf{r}'$. The potential term is moved to the left of the equals sign, which makes the combination $H_i + V_h = H_f$. The auxiliary function $P(\mathbf{r}, i\omega)$ obeys the equation

$$(i\omega - \omega_T - H_f)P(\mathbf{r}, i\omega) = \hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}) \tag{9.136}$$

This equation is solved in terms of the basis functions Ψ_f of the final state Hamiltonian H_f:

$$P(\mathbf{r}, i\omega) = \sum_{\mathbf{k}} \frac{\Psi_f(\mathbf{k}, \mathbf{r})\int d^3r'\Psi_f^*(\mathbf{k}, \mathbf{r}')\hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}')}{i\omega - \omega_T - \varepsilon_f(\mathbf{k})} \tag{9.137}$$

An obvious notation is to define the optical matrix element in terms of the final state basis functions:

$$w_f(\mathbf{k}) = \int d^3r\Psi_f(\mathbf{k}, \mathbf{r})\hat{\varepsilon}\cdot\mathbf{p}\phi_c(\mathbf{r}) \tag{9.138}$$

The correlation function in (9.131) is

$$\pi(i\omega) = \frac{1}{\nu}\sum_{\mathbf{k}s}\frac{|w_f(\mathbf{k})|^2}{i\omega - \omega_T - \varepsilon_f(\mathbf{k})} \tag{9.139}$$

$$A(\omega) = \frac{2\pi}{\nu}\sum_{\mathbf{k}s}|w_f(\mathbf{k})|^2\delta[\omega - \omega_T - \varepsilon_f(\mathbf{k})] \tag{9.140}$$

The absorption spectrum is the spectral function of this correlation function. Equation (9.140) is actually a simple physical result. It shows that the absorption spectrum $A(\omega)$ can be obtained from a one-particle spectrum by using the golden rule. However, one has to use the wave functions $\Psi_f(\mathbf{k}, \mathbf{r})$ and eigenvalues ε_f of the electron in the final state of the transition. During the optical transition, the effective potential of the system changes from V_i to V_f. Formula (9.140) shows that the optical spectrum is calculated using V_f. This conclusion is important, not only in solid-state spectra but also in atomic physics.

In a single-particle picture, this conclusion is obvious. The core hole wave function $\phi_c(\mathbf{r})$ is also calculated for the potential V_f, since in calculating such states one does not include the interaction of a particle with itself. That is, presumedly ϕ_c is found from a Schrödinger equation with a potential which also does not have the core level occupied. In a single-particle picture, this potential is V_f. The final formula (9.140) merely describes a one-particle transition between two different energy levels of the same Hamiltonian H_f. It is an unremarkable application of the golden rule for transition rates. It is possible to solve this many-body problem because it is a one-body problem.

Unfortunately the present model is too simple to apply accurately to real spectra. One of the assumptions is that the changeover from V_i to V_f occurs instantly with the creation of the hole. In actual systems, the other electrons and ions take time to adjust to the core hole potential. This time can be viewed as a frequency-dependent screening process. A realistic calculation should confront the dynamic screening, which was carefully neglected at the beginning of the calculation.

9.3. X-RAY SPECTRA IN METALS

9.3.1. Physical Model

The previous section discussed exciton effects in the core-level spectra of semiconductors or insulators. Now consider the same optical transitions in metals. The photons usually have higher energy, so the spectroscopy is labeled X-ray or soft X-ray. The many-body theory of this spectroscopy is considerably more complicated, since the conduction electrons in the metal respond dramatically to the X-ray transition. These processes cause several new effects which were predicted theoretically and verified experimentally. Among them are (1) the prediction of Mahan (1967) that the absorption edges had a power law divergence near threshold, which has been found for the p-shell spectra of simple metals such as sodium, magnesium, and potassium, and (2) the prediction of Doniach and Sunjic (1970) of asymmetries in the XPS (X-ray photoelectron spectroscopy) line shapes from core levels of these metals, which is also now well documented experimentally. The latter phenomenon is a consequence of (3) the orthogonality catastrophe, first explained by Anderson (1967). The general problem is often called "MND" after the work of Mahan (1967) and Nozieres and deDominicis (1969).

The model for free-electron metals is shown in Fig. 9.9. It is an energy-level diagram and shows the parabolic states in the conduction band, which is filled with electrons to the Fermi energy μ. The horizontal lines on the bottom represent core levels in the ions in the metal; they are localized, have an infinite effective mass, and are drawn flat on an energy-level diagram. In the X-ray absorption process, the photon energy is used to lift an electron from a core level to an unoccupied state in the conduction band. At zero temperature, the only empty states are above the chemical potential μ. In a one-election description of this process, the absorption must start at the threshold frequency $\hbar\omega_T = E_F + E_c$ where E_c is the core-level binding energy, as measured from the bottom of the conduction band, and E_F is the width of the occupied conduction band:

$$A^{(0)}(\omega) = 2\pi \int \frac{d^3k}{(2\pi)^3} |\langle \mathbf{k}|\hat{\varepsilon}\cdot\mathbf{p}|c\rangle|^2 [1 - n_F(\xi_k)]\delta(\hbar\omega - \hbar\omega_T - \xi_k) \tag{9.141}$$

The factor $[1 - n_F]$ is included to limit the final states to above the Fermi level. As indicated schematically in Fig. 9.9, the golden rule result (9.141) predicts that the absorption starts with a nonzero value at threshold, so there is a step in the spectrum. The step occurs because the electron gas has a nonzero density of states at the threshold frequency, since the electrons are going to states right above the Fermi energy.

The core levels of ions have angular momentum as a good quantum number. The angular momentum of the electron state must change by one unit during the photon transition. If the core level has an initial angular momentum l, the final state in the conduction band is either $l+1$ or $l-1$. An atomic description of this absorption process would write the one-particle theory in terms of these two final state channels:

$$A^{(0)}(\omega) = 2\pi \int \frac{d^3k}{(2\pi)^3} [A_{l+1}(k) + A_{l-1}(k)][1 - n_F(\xi_k)]\delta(\hbar\omega - \hbar\omega_T - \xi_k) \tag{9.142}$$

where the constants $A_{l\pm1}(k)$ are found from Clebsch–Gordan coefficients and radial integrals. In a metal, the final electron state is a Bloch function, which is not an eigenfunction of the angular momentum operator. The use of Bloch functions leads to a more complicated

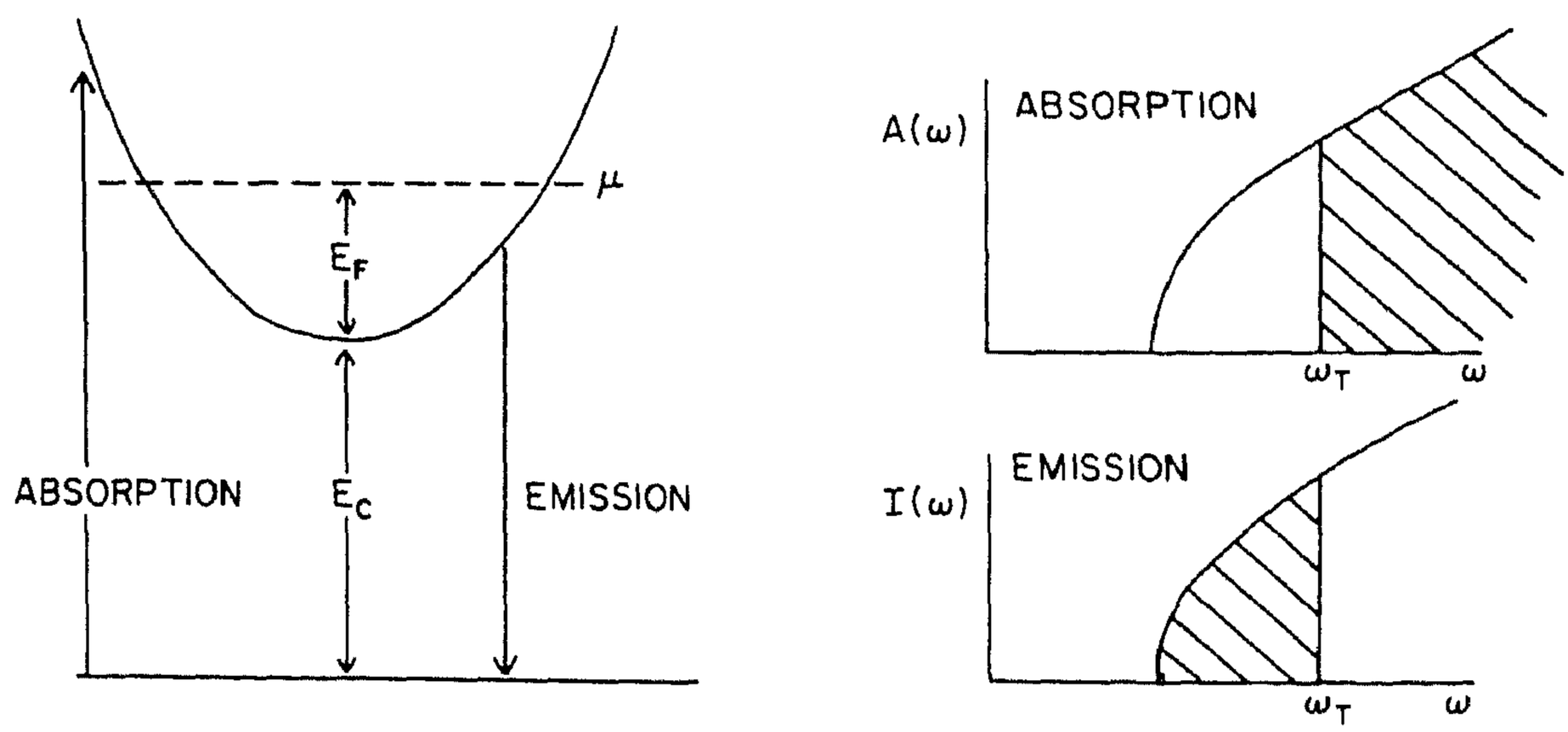


FIGURE 9.9 X-ray transition between core levels and parabolic conduction band of a metal. Electron sates are occupied below Fermi energy which is E_F above the conduction band minimum. Absorption must remove a core electron and out it into an unoccupied band state, while emission must remove an electron from an occupied band state. Also shown are the predicted absorption and emission spectra in the one-body picture, neglecting energy band variations.

description, since the final electron states in (9.141) must be summed over the full Brillouin zone and its various energy bands.

The word *hole* has two different meanings in the present discussion. The first is a core hole, which is an electron missing from a core state. The other is a conduction band hole, where the electron is absent from an occupied state in the Fermi sea. Core hole is used for the first case and hole for the second.

After the core hole has been created, it has several possible decay channels. The most likely is an Auger process, whereby an electron from a higher energy state falls into this core hole while its energy is transferred to another electron whose energy is correspondingly increased. The electron which gains energy can also be measured, and this Auger spectroscopy is another important experiment (Gallon, 1978).

The second decay channel is the emission of a photon by an electron which falls into the core hole. When the electron starts in the occupied states in the conduction band, this X-ray emission spectroscopy is an experiment which is complementary to the absorption. The emission provides a measurement over the states, below the Fermi level, which are occupied in the ground state of the Fermi sea. The absorption measures the unoccupied states above the Fermi energy. The emission spectrum also has a sharp step at the threshold frequency $\hbar\omega_T = E_F + E_c$, as shown schematically in Fig. 9.9. These steps, in emission and absorption, are called *Fermi edges*.

There are several physical mechanisms which can change the shape of the spectrum, in emission or absorption, from the simple one-particle theory in (9.141). The first of these is due to the lifetime of the core hole, which is usually dominated by the Auger decay process as just described. This core hole lifetime is quite variable among the various core levels of an ion or atom. The core hole lifetime is longer for the outer valence shells than for the inner states closer to the nucleus. Some experimental values for core hole lifetimes in simple metals are shown in Table 9.1, from data of Citrin *et al.* (1977). The same group recently remeasured these values, Wertheim *et al.* (1992). The Auger lifetime is expressed as an energy uncertainty $\Gamma = \hbar/\tau$ (where Γ is the full width at half maximum). Values as high as several electron volts can be found for inner shells of atoms with high Atomic number. This Auger decay generally imparts a Lorentzian broadening to the X-ray spectra. It can be included, semiempirically, in the Kubo formula as a factor $\exp(-\Gamma|t|/2)$. If the X-ray measurement is to determine other phenomena, it is desirable to keep Γ, as small as possible, which is done by measuring the spectra of outer atomic shells.

TABLE 9.1 Core level data at $T = 300$ K. Γ is Auger width while γ is phonon width. From Citrin *et al.* (1977)

Metal	State	Γ(eV)	γ(eV)	α
Li	$1s$	0.03	0.37	0.23
Na	$2p$	0.02	0.18	0.20
	$2s$	0.28	0.20	0.20
	$1s$	0.28	0.18	0.21
Mg	$2p$	0.03	0.18	0.13
	$2s$	0.46	0.19	0.13
	$1s$	0.35	0.20	0.15
Al	$2p$	0.04	0.11	0.12
	$2s$	0.78	—	0.12

Another many-body effect is the broadening due to phonons. The only term which is usually kept in the Hamiltonian is the phonon coupling to the core hole, which is a term such as

$$V = d^\dagger d \sum_{\mathbf{q}\lambda} M_\lambda(\mathbf{q})(a_{\mathbf{q}\lambda} + a^\dagger_{-\mathbf{q}\lambda}) \tag{9.143}$$

It was shown in Sec. 4.3 that this type of phonon coupling to a localized level is a type of Hamiltonian which is exactly solvable. The optical absorption from such a coupled level was also solved, and the Kubo formula has a factor of $\exp[-\phi(t)]$, where

$$\begin{aligned}\phi(t) &= \sum_{\mathbf{p}\lambda} \left|\frac{M_\lambda(\mathbf{q})}{\omega_\lambda(\mathbf{q})}\right|^2 \{[1 - it\omega_\lambda(\mathbf{q}) - e^{-it\omega_\lambda(\mathbf{q})}][n_B(\omega_\lambda(\mathbf{q})) + 1] \\ &\quad + [1 + it\omega_\lambda(\mathbf{q}) - e^{it\omega_\lambda(\mathbf{q})}]n_B(\omega_\lambda(\mathbf{q}))\} \qquad (9.144)\\ &\approx \gamma^2 t^2, \qquad \gamma^2 = \tfrac{1}{2}\sum_{\mathbf{q}\lambda} |M_\lambda(\mathbf{q})|^2[1 + 2n_B(\omega_\lambda(\mathbf{q})] \qquad (9.145)\end{aligned}$$

The term linear in t is the self-energy term. The factor $\exp(-\phi)$ is added to the argument of the time integral in (8.3.3). Usually the Auger width Γ is large enough that only the short time response is needed of the correlation function. Anything that is going to happen must do so quickly, before the core hole decays by the Auger process. The short time limit of the phonon contribution is $\phi \approx \gamma^2 t^2$ as shown above. The spectral function is broadened by the Auger and phonon contributions

$$A(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} \langle \mathscr{J}(t)\mathscr{J}(0)\rangle \exp(-\tfrac{1}{2}\Gamma|t| - \gamma^2 t^2) \tag{9.146}$$

The phonons add a Gaussian broadening to the spectral function for the X-ray process, where the Gaussian component γ is temperature dependent. It is relatively easy to experimentally separate the phonons from the Auger width, which is a temperature-independent Lorentzian. The only remaining problem for the phonon contribution is to calculate the matrix elements $M_\lambda(\mathbf{q})$ for the coupling to the core hole. However, this calculation is quite delicate (Hedin and Rosengren, 1977).

The phonon broadening for the emission spectra is a much more subtle calculation. Here the problem is that the phonon system does not equilibrate to the presence of the core hole before the emission occurs. One expects that the minimum time for the phonons to come to equilibrium around the new core hole, and its screening charge, is the inverse Debye frequency $t_{ph} \approx 1/\omega_D$. This time is generally much longer than the lifetime for the Auger effect, except for the outer shell of electrons, as shown in Table 9.1. The theory of phonon broadening of emission must account for the nonequilibrium states of the phonons. This theory was developed simultaneously by Mahan (1977) and Almbladh (1977). The most interesting case yet known is lithium, for which phonons are the most important of all the mechanisms which broaden the edge. The theory predicts a double shoulder in the emission spectra because of incomplete relaxation of the phonons. The prediction was verified by the experiments of Callcott and Arakawa (1977).

These many-body processes, Auger and phonon, are reasonably straightforward. The more interesting aspects are the response of the many-electron system to the appearance of the core hole, which are discussed in the remaining parts of this section.

9.3.2. Edge Singularities

Mahan (1967) predicted that the absorption edges for X-ray transitions in metals would have a power law divergence of the form

$$A(\omega) = A^{(0)}(\omega)\left(\frac{\xi_0}{\omega - \omega_T}\right)^{\alpha} \tag{9.147}$$

where $\xi_0 \sim E_F$ is a bandwidth, and the exponent α is discussed below. His theory resulted from an investigation into the properties of excitons in metals. The Wannier picture applies equally well to metals, so that the optical transition should be viewed as the simultaneous creation of an electron and a core hole in the metal. These two particles will interact in the final state. Presumably this interaction is screened, although a detailed theory of the screening process is still an active subject of research (Canright, 1988). Mahan assumed that the electron core hole Coulomb potential is screened instantly at the time the core hole is created. This approximation may be valid near the absorption edges but surely is unreliable farther from threshold.

The exciton theory predicts that the exponent α in (9.147) is positive, so the absorption diverges as a power law singularity at threshold. Later work has shown that the renormalization catastrophe of Anderson tends to make α negative. The final values of α may be either positive or negative, depending on which of these two factors is most important. Not all edges are singular.

The calculation assumes there is instantaneous screening of the electron–hole interaction in the metal. The model Hamiltonian for the X-ray absorption process is:

$$H = \xi_h d^\dagger d + \sum_{\mathbf{k}} \xi_c(\mathbf{k}) C_{\mathbf{k}}^\dagger C_{\mathbf{k}} + \frac{1}{\nu}\sum_{\mathbf{k}\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') C_{\mathbf{k}}^\dagger C_{\mathbf{k}'} d^\dagger d \tag{9.148}$$

The spin indices have been dropped from the notation. The last term is the screened Coulomb interaction between the electron and the core hole, while the first terms are H_0 for the electrons and the core holes. The Hamiltonian is written in the initial state basis, so the electron operators $(C_{\mathbf{k}}, C_{\mathbf{k}}^\dagger)$ refer to the state which does not have a core hole. For the Wannier exciton theory it was shown that there is a great advantage to working in the final state basis set, but that is not the case here. The calculation is about equally difficult in either basis set, and the same answer is found from either starting point. The difference between the present calculation for the metal, and the earlier calculation for the insulator, is that there are now $N \sim 10^{23}$ conduction electrons which also interact with the core hole, and they are initially in states described by the initial state basis. In the calculation, it is now necessary to keep the terms $n_F(\xi_c)$ for the electron occupation number. They are the cause of the dramatic change in the theory, and the prediction (9.147) of the edge singularity.

In the exciton calculation, the vertex diagrams to the correlation function are included by summing the set of diagrams in Fig. 9.10. To keep the physics simple, the calculation is first done by assuming some approximations which tend to reduce the number of superscripts and subscripts: (1) the interband matrix element $\langle f|\hat{\varepsilon}\cdot p|c\rangle$ is taken to be a constant w_0; (2) the screened Coulomb interaction is taken to be a constant $V(\mathbf{k}, \mathbf{k}') = V_0$ up to a cutoff energy $\xi_0 \approx E_F$ which is a typical bandwidth. With these approximations, the ladder diagrams in the

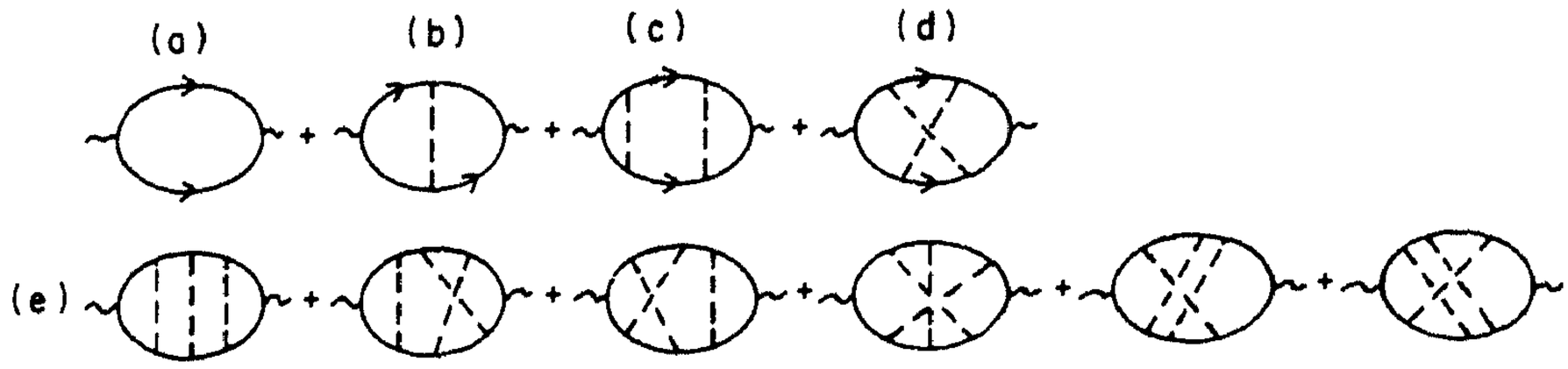


FIGURE 9.10

vertex summation in Fig. 9.10 can be written as in the previous section:

$$\pi_L(i\omega) = \sum_{n=0} \pi^{(n)}(i\omega) \tag{9.149}$$

$$\pi^{(n)}(i\omega) = w_0^2(-V_0)^n \left[\int \frac{d^3k}{(2\pi)^3} \frac{1 - n_F(\xi_k) - n_F(\xi_h)}{i\omega - \omega_T - \xi_{\mathbf{k}}} \right]^{n+1} \tag{9.150}$$

Removing the wave vector dependence of the scattering potential uncouples the wave vector integrals. One is left with a product of $(n + 1)$ integrals which are identical. The integrand of the wave vector integral contains the factor $1 - n_F(\xi_k) - n_F(\xi_h)$, where the core hole term $n_F(\xi_h)$ is set equal to zero. This step leaves $1 - n_F(\xi_{\mathbf{k}})$, which ensures that the electron scatters into unoccupied states. The summation over wave vectors can be changed approximately to $d^3k \to (2\pi)^3 N_F d\xi_{\mathbf{k}'}$, where N_F is the density of states. The $d\xi_{\mathbf{k}'}$ integral is just a logarithm, and the ladder diagrams give the result at zero temperature

$$N_F \int_{-\xi_0}^{\xi_0} d\xi_{\mathbf{k}'} \frac{1 - n_F(\xi_{\mathbf{k}})}{i\omega - \omega_T - \xi_{\mathbf{k}}} = N_F \ln\left(\frac{\omega_T - i\omega}{\xi_0}\right) \tag{9.151}$$

$$\pi^{(n)}(i\omega) = -w_0^2 N_F(-N_F V_0)^n \left[\ln\left(\frac{\omega_T - i\omega}{\xi_0}\right) \right]^{n+1} \tag{9.152}$$

These approximations are fairly crude. They are meant only to apply to the threshold region, where $\omega \approx \omega_T$, so that the cutoff in the wave vector integration by the Fermi occupation factors n_F is most important. The vertex summation leads to a summation over logarithms of increasing powers.

This series should not yet be summed, since other vertex diagrams contribute terms which also contribute to the series. The first diagram in the series, which is not a ladder, is that shown in Fig. 9.10(d). It contains six Green's functions as internal lines. The following summations over Matsubara frequencies must be done:

$$\begin{aligned} &\frac{1}{\beta^3} \sum_{ik,iq,iq'} \mathcal{G}_e(\mathbf{k}, ik)\mathcal{G}_e(\mathbf{k}+\mathbf{q}, ik+iq)\mathcal{G}_e(\mathbf{k}+\mathbf{q}+\mathbf{q}', ik+iq+iq') \\ &\quad \times \mathcal{G}_h(-\mathbf{k}, i\omega - ik)\mathcal{G}_h(-\mathbf{k}-\mathbf{q}', i\omega - ik - iq') \\ &\quad \times \mathcal{G}_h(-\mathbf{k}-\mathbf{q}-\mathbf{q}', i\omega - ik - iq - iq') \end{aligned} \tag{9.153}$$

These summations are done in sequence. The core hole Green's functions all have the same energy ξ_h which is independent of wave vector. The conduction electron energies will be denoted by the symbols $\xi = \xi_e(\mathbf{k})$, $\xi' = \xi_e(\mathbf{k}+\mathbf{q})$, $\xi'' = \xi_e(\mathbf{k}+\mathbf{q}+\mathbf{q}')$. When doing these

summations, the core hole occupation factor $n_F(\xi_h)$ is set to zero wherever it occurs. With these conventions, the three summations are, in sequence,

$$\frac{1}{\beta}\sum_{iq'} \mathcal{G}_e(k+q+q')\mathcal{G}_h(i\omega - ik - iq')\mathcal{G}_h(i\omega - ik - iq - iq')$$

$$= \frac{1 - n_F(\xi'')}{(i\omega - \xi'' - \xi_h)(i\omega + iq - \xi'' - \xi_h)}$$

$$\frac{1}{\beta}\sum_{iq} \frac{\mathcal{G}_e(k+q)}{i\omega + iq - \xi'' - \xi_h} = \frac{n_F(\xi')}{i\omega - ik + \xi' - \xi'' - \xi_h}$$

$$\frac{1}{\beta}\sum_{ik} \frac{\mathcal{G}_e(k)\mathcal{G}_h(i\omega - ik)}{i\omega - ik + \xi' - \xi'' - \xi_h} = \frac{1 - n_F(\xi)}{(i\omega - \xi - \xi_h)(i\omega - \xi + \xi' - \xi'' - \xi_h)}$$

The vertex correction is obtained from the integral ($\xi_h = \omega_T$)

$$\pi^{(2b)}(i\omega) = V_0^2 N_F^3 \omega_0^2 \int_{-\xi_0}^{\xi_0} d\xi \int_{-\xi_0}^{\xi_0} d\xi' \int_{-\xi_0}^{\xi_0} d\xi'' \times \frac{[1 - n_F(\xi)]n_F(\xi')[1 - n_F(\xi'')]}{(i\omega - \xi - \omega_T)(i\omega - \xi + \xi' - \xi'' - \omega_T)(i\omega - \xi'' - \omega_T)} \tag{9.154}$$

This integral is more difficult to evaluate than those in the previous cases, since the middle energy denominator contains all three conduction energies, $-\xi + \xi' - \xi''$. The integral over $d\xi'$ is over the occupied states $n_F(\xi')$. These terms could be ignored in the insulator where n_0 is zero. They must be included in the metal where $n_0 \neq 0$.

The triple integral in (9.154) is obviously complicated. The most divergent term it yields is

$$\pi^{(2b)}(i\omega) = \tfrac{1}{3} w_0^2 N_F (N_F V_0)^2 \left[\ln\left(\frac{\omega_T - i\omega}{\xi_0}\right)\right]^3 \tag{9.155}$$

This term is one-third the contribution of the double ladder diagram in Fig. 9.10(c) and with the opposite sign.

A similar calculation can be done for the third order vertex corrections. Each of the six terms with three vertex lines in Fig. 9.10(e) produces terms which have $\{\ln[(\omega_T - i\omega)/\xi_0]\}^3$. The summation of all the terms shown in Fig. 9.10 gives the series

$$\pi(i\omega) = -\frac{N_F w_0^2}{N_F V_0}(L - L^2 + \tfrac{2}{3}L^3 - \tfrac{1}{3}L^4 + \cdots) \tag{9.156}$$

$$L = N_F V_0 \ln\left(\frac{\omega_T - i\omega}{\xi_0}\right) \tag{9.157}$$

The series is now summed to produce the final result:

$$\pi(i\omega) = -\frac{N_F w_0^2}{2N_F V_0}(1 - e^{-2L}) \tag{9.158}$$

$$= -\frac{N_F w_0^2}{2N_F V_0}\left[1 - \left(\frac{\xi_0}{\omega_T - i\omega)}\right)^{2N_F V_0}\right] \tag{9.159}$$

The optical absorption is the spectral function of this operator. Note that the right-hand side is real for $\omega < \omega_T$, but complex parts are generated when $\omega > \omega_T$

$$\begin{aligned} A(\omega) &= -2\,\mathrm{Im}[\pi(\omega)] \\ &= -\frac{N_F w_0^2}{N_F V_0}\mathrm{Im}\left\{\left(\frac{\xi_0}{\omega-\omega_T}\right)^{2N_F V_0}\exp[-i2\pi N_F V_0\Theta(\omega-\omega_T)]\right\} \\ &= \frac{w_0^2}{V_0}\sin(2\pi N_F V_0)\left(\frac{\xi_0}{\omega-\omega_T)}\right)^{2N_F V_0}\Theta(\omega-\omega_T) \end{aligned} \tag{9.160}$$

Equation (9.160) is Mahan's (1967) result that the absorption diverges as a power law, where the exponent is the dimensionless quantity $\alpha = 2N_F V_0$. The exponent is obtained by summing the most divergent diagrams in each order of perturbation theory. These terms are divergent, because the retarded function $\ln[(\omega_T - \omega)/\xi_0]$ can become quite large near threshold $\omega \approx \omega_T$. In the limit where $V_0 = 0$, the above formula goes to the noninteracting result that $A^{(0)} = 2\pi N_F w_0^2 \Theta(\omega - \omega_T)$, and the absorption edge is a simple step.

The correlation function (9.77) for the model Hamiltonian (9.148) can be evaluated accurately using simple analytical formulas. Pardee and Mahan (1973) showed that the edge singularities could be derived in a simple way using dispersion theory. Penn *et al.* (1981, called PGM) obtained an analytical solution to the multiple scattering problem, including all vertex corrections. They showed that the leading term in the series was the result of Pardee and Mahan. Ohtaka and Tanabe (1983, 1986) obtained a more elegant analytical solution, which they illustrated with numerical examples. This work has provided a complete solution to the MND problem, and in an analytical form that is useful for computation.

The result of Pardee and Mahan is obtained by a simple argument. Let $T(\omega)$ be the complex scattering amplitude for an ingoing wave of an electron scattering from a central potential, which in this case is the screened core hole. If $T'(\omega)$ is the amplitude of the outgoing scattered wave, they differ only by a phase factor which is twice the phase shift:

$$T'(\omega) = T(\omega)\exp[2i\delta_l(\omega)\Theta(\omega-\omega_T)] \tag{9.161}$$

where l is the angular momentum of the outgoing electron. The step function is a reminder that the phase factor exists only in the region of absorption, since T and T' must both be real in regions where there is no absorption. The next assumption is that T and T' both originate from the same analytical function of frequency:

$$t(\omega + i\delta) = T'(\omega) \tag{9.162}$$

$$t(\omega - i\delta) = T(\omega) \tag{9.163}$$

Combining these results shows that

$$t(\omega \pm i\delta) = \exp[\phi(\omega \pm i\delta)] = |T(\omega)|\exp[\pm i\delta_l(\omega)\Theta(\omega-\omega_T)] \tag{9.164}$$

$$\mathrm{Im}[\phi(\omega \pm i\delta)] = \pm\delta_l(\omega)\Theta(\omega-\omega_T) \tag{9.165}$$

Dispersion theory assumes that $\phi(z)$ in (9.164) is an analytical function of complex frequency z. The only function that has the property (9.164) is

$$\phi(z) = \frac{1}{\pi}\int_{\omega_T}^{\xi_0+\omega_T} d\varepsilon\,\frac{\delta_l(\varepsilon)}{\varepsilon - z} \tag{9.166}$$

As usual ξ_0 is the bandwidth. The function $\phi(z)$ is a dispersion integral. The integral can be evaluated by writing $\delta_l(\varepsilon) = \delta_l(\omega) + [\delta_l(\varepsilon) - \delta_l(\omega)]$ and the first term is an easy integral:

$$\phi(\omega + i\delta) = \frac{\delta_l(\omega)}{\pi}\left[\ln\left|\frac{\xi_0 + \omega_T - \omega}{\omega_T - \omega}\right| + i\pi\Theta(\omega - \omega_T)\right] + K(\omega) \tag{9.167}$$

$$K(\omega) = \frac{1}{\pi}\int_{\omega_T}^{\xi_0} d\varepsilon \frac{\delta_l(\varepsilon) - \delta_l(\omega)}{\varepsilon - \omega} \tag{9.168}$$

The function $K(\omega)$ is a smooth function of ω since the integrand is real and not singular. The amplitude of the X-ray absorption is proportional to $T'(\omega)$ and the intensity is proportional to $|T'(\omega)|^2$. This factor can be evaluated using the dispersion integrals

$$|T'(\omega)|^2 = \left|\frac{\xi_0 + \omega_T - \omega}{\omega_T - \omega}\right|^{2\delta_l(\omega)/\pi} e^{2K(\omega)} \tag{9.169}$$

$$A(\omega) \approx \frac{w_0^2}{V_0}\sin(2\delta)\left(\frac{\xi_0}{\omega - \omega_T}\right)^{2\delta/\pi} \Theta(\omega - \omega_T) \tag{9.170}$$

This equation has the same edge singularity as (9.160). Here the exponent of the power law is $\alpha = 2\delta_l(\omega)/\pi$ rather than $2N_F V_0$. Note that if one is calculating the phase shift in the Born approximation, then one gets that $\sin(\delta) \approx \delta \approx \pi N_F V_0$. Nozieres and deDominicis (1969) were the first to realize that the exponents of the edge singularities were functions of the phase shifts. At the threshold frequency $\omega \approx \omega_T$ then the exponent depends upon $\delta_l(\omega_T)$, which are the phase shifts of conduction electrons at the Fermi surface for scattering from the core hole. The factor of $K(\omega)$ is a smooth function of frequency and does not contribute to the singular behavior. It has an influence upon the calculation of the X-ray absorption for other values of frequency.

In X-ray absorption an electron absorbs an X-ray, departs the atomic core, and becomes a conduction electron. The core potential changes from having a charge of Z to having a charge of $Z + 1$. All of the other conduction electrons have their central potential suddenly changed by the appearance of this new core hole. The switching-on of the potential may cause electron–hole pairs to be created. The pairs act as bosons. The theory for this process can be treated by the independent Boson model of Sec. 4.3.

The electron that departs the core state and becomes a conduction electron may also create excitations of electron–hole pairs. The departing electron has only a small probability of creating pairs. This process may be accurately calculated by perturbation theory.

X-ray absorption has a number of factors that contribute to the shape of the spectra. The above discussion derived the contributions of phonons, Auger decay, and the final state Coulomb interactions. Another important contribution is the shakeup mentioned above. The shakeup contributes a time evolution $\rho(t)$ which is calculated in the next section. All of these various factors contribute a time dependence to the evolution of the absorption. The experimental spectra are found by Fourier transforming the product of all of these factors

$$A(\omega) = \int_{-\infty}^{\infty} dt\rho(t)A(t)\exp[it(\omega - \omega_T) - \Gamma|t| - \gamma^2 t^2] \tag{9.171}$$

The factor $A(t)$ is the Fourier transform of $A(\omega)$ in (9.170). Instead of transforming $A(\omega)$, an alternative is to represent (9.171) as a convolution integral. The evaluation of $\rho(t)$ is presented in the next section.

9.3.3. Orthogonality Catastrophe

The X-ray absorption creates a core hole in the midst of the electron gas. It also creates an additional conduction electron, but the latter process has been discussed in the previous section. Now consider the impact on the $N \approx 10^{23}$ other conduction electrons of this sudden appearance of a core hole. The core hole is again represented as a static potential V. The physics is to investigate the response of the free-electron gas to the sudden appearance of a new potential V. The result has been labeled a *catastrophe* because the transition is forbidden. The word *forbidden* is used in an unusual way, which is now explained.

The important physics is that the ground state wave functions of the conduction electron system, with and without the core hole potential V, are orthogonal. The transition is forbidden because it is the transition between two orthogonal states. The X-ray absorption is observed in actual metals, because the conduction electron system is excited by the creation of electron–hole pairs. The pair creation is a symmetry-breaking process, which allows the forbidden transition to be observed. The orthogonality was first suggested by Hopfield (1967, 1969) and proved by Anderson (1967).

It is important to realize that the orthogonality between the two wave functions, with and without the core hole potential, is between the N-particle wave functions in the limit where $N \to \infty$. The single-particle wave functions are not orthogonal. As a simple example, assume that in the ground state of the electron gas, before the appearance of the core hole, the conduction electrons can be described by single-particle wave functions with no potential, such as plane waves. The derivation follows Anderson who considered only s waves, so that the wave functions have the form $\phi(kr) = \sin(kr)/kr$. In the Hartree–Fock approximation, the N-particle wave function can be described by an N-dimensional Slater determinant of these orbitals. The spin indices are omitted, so write the initial state of the system as

$$\Phi_i(r_1, r_2 \cdots r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi(k_1 r_1) & \phi(k_1 r_2) & \cdots & \phi(k_1 r_N) \\ \phi(k_2 r_1) & \phi(k_2 r_2) & \cdots & \phi(k_2 r_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(k_N r_1) & \phi(k_N r_2) & \cdots & \phi(k_N r_N) \end{vmatrix} \tag{9.172}$$

After the appearance of the core hole potential, the conduction electrons adjust their individual wave functions to the presence of this potential. Far from the potential region, the wave functions can be described by $\phi'(kr) = \sin(kr + \delta)/kr$, where δ is the s-wave phase shift for one-electron scattering from the core hole potential. The new N-particle wave function is just a Slater determinant with these new orbitals ϕ':

$$\Phi_f(r_1, r_2 \cdots r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi'(k_1 r_1) & \phi'(k_1 r_2) & \cdots & \phi'(k_1 r_N) \\ \phi'(k_2 r_1) & \phi'(k_2 r_2) & \cdots & \phi'(k_2 r_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi'(k_N r_1) & \phi'(k_N r_2) & \cdots & \phi'(k_N r_N) \end{vmatrix} \tag{9.173}$$

During the X-ray transition, one takes the matrix element between the initial and final states of the system. In a one-particle picture, the matrix element $\int d^3r \phi_f^*(k_1 r)\hat{\varepsilon} \cdot \mathbf{p}\phi_c(r)$ is between the core wave function $\phi_c(r)$ and a conduction electron $\phi_f(kr)$. The many-particle calculation

computes this matrix element between the N-particle states. It should include the overlap between these two Slater determinants,

$$S = \int d^3r_1 \cdots d^3r_N \Phi_f^*(r_1 \cdots r_N)\Phi_i(r_1 \cdots r_N) \tag{9.174}$$

which has been shown to be

$$S = N^{-\alpha/2} \tag{9.175}$$

$$\alpha = 2\frac{\delta^2}{\pi^2} \tag{9.176}$$

The many-particle overlap S turns out to be a negative power of N, where the exponent is a function of the phase shifts. This overlap vanishes in the limit where $N \to \infty$ and is small for real systems which have $N \sim 10^{23}$ and α is typically 0.1–0.2.

There is no orthogonality between the individual matrix elements of single-particle orbitals $\phi(kr)$ and $\phi'(kr)$. This single integral is nonzero. When these overlaps are evaluated in the determinant for all the possible combinations, the result is asymptotically zero as $N \to \infty$. The analogous situation occurs in atoms but on a less drastic scale. For an atom with N electrons, the optical absorption can cause one electron to change from one energy state to another. The matrix element for this process is calculated for the full N-particle wave function of the atom. The N-particle matrix element can be reduced to a one-particle matrix element between the two primary states which are the initial and final levels of the electron, times a factor S which gives the overlap of the other $N-1$ electrons. They each change their orbitals a small amount during this change of state by one electron, and the $N-1$ overlap function has a typical value of 0.95. Then S is near unity, and the N-particle overlap is a small effect in atoms with a few electrons. The metal case is the extrapolation to $N \sim 10^{23}$ electrons, where the product of one-particle overlaps, each slightly less than unity, eventually produces a vanishing matrix element.

It is necessary to turn these ideas into a dynamical theory of the absorption process. Nozieres and deDominicis (1969) started by examining the Green's function for the core hole:

$$G_h(t) = -i\langle Td(t)d^\dagger(0)\rangle = -i\Theta(t)\langle|e^{iHt}de^{-iHt}d^\dagger|\rangle \tag{9.177}$$

They manipulate this correlation function to show that it can be treated as the many-electron response to the sudden switching on of the core hole, in the same way that the phonon response to sudden switching was studied in Sec. 4.3.5. The core hole creation operator $d^\dagger|\rangle$ acts on the ground state $|\rangle$ to create a single core hole. The operator $\exp(-itH)$ on this state is represented by the final state potential with the core hole:

$$e^{-itH}d^\dagger|\rangle = e^{-it(\omega_T+H_f)t}d^\dagger|\rangle \tag{9.178}$$

The core hole destruction operator d destroys the core hole state. The last time operator $\exp(itH)$ operates on a configuration with no core holes, so that it is H_g for the ground state. The hole core Green's function can be exactly given by

$$G_h(t) = -ie^{-it\omega_T}\Theta(t)\langle|e^{itH_g}e^{-itH_f}|\rangle \tag{9.179}$$

The core hole operators $(d, d^\dagger)$ are now omitted because they have done their work of determining the proper order of the Hamiltonian operators. There are several different ways to evaluate the time dependence of this function. One way is to recognize that the Hamiltonian $H_f = \sum_j h_j$ is a summation of the single-particle Hamiltonians for the individual conduction

electrons. The hole Green's function can be expressed as a Slater determinant analogous to (9.166) for the absorption rate. In fact, it is just the function $\rho(t)$ in (9.169) which is the matrix element over the ground state orbitals of $\exp(-ith_j)$

$$\rho(t) = \langle | e^{itH_g} e^{-itH_f} | \rangle = e^{itE_g} \det[\phi_{\mathbf{pp}'}(t)]_{\mathbf{p},\mathbf{p}'<k_F} \tag{9.180}$$

$$G_h(t) = -i\Theta(t) e^{-it\omega_T} \rho(t) \tag{9.181}$$

$$\phi_{\mathbf{pp}'}(t) = \langle \mathbf{p} | e^{-ith_j} | \mathbf{p}' \rangle \tag{9.182}$$

The evaluation of this determinant is one possible way to find the core hole Green's function. Methods of numerical evaluation have been discussed by Schönhammer and Gunnarsson (1978).

A second way to find the core hole Green's function is by a linked cluster expansion. The correlation function in (9.181) may be written as a time-ordered operator, which is evaluated by the techniques described in Sec. 3.6:

$$\rho(t) = \left\langle T \exp\left[-i \int_0^t dt_1 V(t_1) \right] \right\rangle = \exp\left[\sum_{l=1}^{\infty} F_l(t) \right] \tag{9.183}$$

$$F_l(t) = \frac{(-1)^l}{l} \int_0^t dt_1 \cdots \int_0^t dt_l \langle T V(t_1) \cdots V(t_l) \rangle_{\text{connected}} \tag{9.184}$$

The first term in the exponential resummation $F_1(t)$ is linear in t, and is a self-energy term:

$$F_1(t) = -i \int_0^t dt_1 \langle V(t_1) \rangle = \frac{-it}{\nu} \sum_{\mathbf{k}} V(\mathbf{k}, \mathbf{k}) n_F(\xi_{\mathbf{k}}) \tag{9.185}$$

The exact self-energy of the core hole is known from Fumi's theorem in Sec. 4.1.3 to be

$$E_i = -\frac{2\hbar^2}{m} \sum_l (2l+1) \int_0^{k_F} k dk \delta_l(k) \tag{9.186}$$

The factor $V(\mathbf{k}, \mathbf{k})$ is the first term in an expansion which should give Fumi's result E_i when all terms are summed. It seems reasonable to use Fumi's result (9.186) for the self-energy, which includes all the terms linear in t in higher order. The threshold energy for the X-ray transition was previously given as ω_T, which is the value in the one-electron approximation, which neglects the final state interactions. They provide a self-energy E_i which further lowers the threshold by this amount. The new threshold energy in the interacting system is $\tilde{\omega}_T = \omega_T + E_i$, which accounts for the increased binding of the core hole due to the polarization of the electron gas around it.

The second term in the exponential series $F_2(t)$ is the first term which is interesting, since it predicts an orthogonality catastrophe:

$$F_2(t) = \frac{(-i)^2}{2} \int_0^t dt_1 \int_0^t dt_2 \langle | T_t V(t_1) V(t_2) | \rangle \tag{9.187}$$

$$= \frac{(-i)^2}{2\nu^2} \int_0^t dt_1 \int_0^t dt_2 \sum_{\mathbf{k}_1 \mathbf{k}_2 \sigma} \sum_{\mathbf{k}_3 \mathbf{k}_4 \sigma'} V(\mathbf{k}_1, \mathbf{k}_2) V(\mathbf{k}_3, \mathbf{k}_4)$$

$$\times \langle T_t C^\dagger_{\mathbf{k}_1 \sigma}(t_1) C_{k_2 \sigma}(t_1) C^\dagger_{\mathbf{k}_3 \sigma'}(t_2) C_{\mathbf{k}_4 \sigma'}(t_2) \rangle \tag{9.188}$$

The operators in the brackets are paired, according to Wick's theorem, into Green's functions of time. The time integrals are then evaluated directly:

$$F_2(t) = \frac{(-i)^2}{2\nu^2}\int_0^t dt_1 \int_0^t dt_2 \sum_{\mathbf{k}_1\mathbf{k}_2\sigma} V(\mathbf{k}_1, \mathbf{k}_2)V(\mathbf{k}_2, \mathbf{k}_1)$$

$$\times\, G(\mathbf{k}_1, t_1 - t_2)G(\mathbf{k}_2, t_2 - t_1) \tag{9.189}$$

$$= \frac{1}{2\nu^2}\int_0^t dt_1 \int_0^t dt_2 \sum_{\mathbf{k}_1\mathbf{k}_2\sigma} |V(\mathbf{k}_1, \mathbf{k}_2)|^2 e^{i(t_1-t_2)(\xi_2-\xi_1)}$$

$$\times\, [\theta(t_1 - t_2) - n_1][\theta(t_2 - t_1) - n_2] \tag{9.190}$$

$$F_2(t) = \frac{1}{\nu^2}\sum_{\mathbf{k}_1\mathbf{k}_2\sigma} |V(\mathbf{k}_1, \mathbf{k}_2)|^2 \left[\frac{itn_1}{\xi_1 - \xi_2} - \frac{n_1(1-n_2)}{(\xi_1-\xi_2)^2}(1 - e^{it(\xi_1-\xi_2)})\right]$$

The factor of $\frac{1}{2}$ in front vanished, because each term in the final result appears twice. The summation over spin index σ can be taken to produce another factor of 2 in front. The first term, which is linear in time, is dropped because it contributes only to the self-energy, which is already known from Fumi's theorem. Consider the expression

$$F_2(t) = -\frac{2}{\nu^2}\sum_{\mathbf{k}_1\mathbf{k}_2} |V(\mathbf{k}_1, \mathbf{k}_2)|^2 \frac{n_F(\xi_{\mathbf{k}_1})[1 - n_F(\xi_{\mathbf{k}_2})]}{(\xi_{\mathbf{k}_1} - \xi_{\mathbf{k}_2})^2}\left(1 - e^{it(\xi_{\mathbf{k}_1}-\xi_{\mathbf{k}_2})}\right)$$

This term does predict an orthogonality catastrophe and has been investigated extensively; see Minnhagen (1977) for an evaluation. Here the topic is discussed briefly. Rewrite this expression by assuming that the hole potential $V(\mathbf{k}, \mathbf{k}')$ depends only on the difference of its arguments $V(\mathbf{k} - \mathbf{k}')$ and change variables to $\mathbf{k}_2 = \mathbf{k}_1 + \mathbf{q}$:

$$F_2(t) = -\int_0^\infty \frac{du}{u^2} R_e(u)(1 - e^{-iut}) \tag{9.191}$$

$$R_e(u) = \frac{1}{\pi\nu}\sum_{\mathbf{q}} V(\mathbf{q})^2 \Lambda(\mathbf{q}, u) \tag{9.192}$$

$$\Lambda(\mathbf{q}, u) = \frac{2\pi}{\nu}\sum_{\mathbf{k}} n_F(\xi_{\mathbf{k}})[1 - n_F(\xi_{\mathbf{k}+\mathbf{q}})]\delta(u + \xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{q}}) \tag{9.193}$$

The factor $\Lambda(\mathbf{q}, u)$ is recognized as the imaginary part of the polarization diagram $P^{(1)}(\mathbf{q}, u)$ of a single-electron bubble. It was evaluated in Sec. 5.5:

$$\Lambda(\mathbf{q}, u) = -\mathrm{Im}[P^{(1)}(\mathbf{q}, u)] = u\left(\frac{m^2}{2\pi q}\right)\Theta(2k_F - q) \tag{9.194}$$

The important feature of this result is its linear dependence on the frequency u, which is valid at small u. The wave vector integration over q can be done for realistic potentials $V(\mathbf{q})$. A

crude approximation to the result is $2(N_F V_0)^2$ where $N_F = mk_F/2\pi^2$ is the density of single spin states at the Fermi energy, while V_0 is the magnitude of the electron–hole potential.

$$R_e(u) = \begin{cases} gu & u < \xi_0 \\ 0 & u > \xi_0 \end{cases} \tag{9.195}$$

$$g = \frac{m^2}{2\pi^2}\int \frac{d^3q}{(2\pi)^3}\frac{1}{q}V(q)^2\theta(2k_F - q) \approx 2(N_F V_0)^2 \tag{9.196}$$

The upper limit of u is ξ_0, which is determined by the range over which $R_e(u)$ is linear in u. This range is approximately $\xi_0 \sim E_F$, as shown in Sec. 5.5:

$$F_2(t) = g\int_0^{\xi_0} \frac{du}{u}(1 - e^{-iut}) \tag{9.197}$$

The remaining du integral is interesting. The integrand has the factor $\int_0 du/u$ which could diverge as ln(0). In this case it does not diverge, since the other factor $(1 - e^{-iut}) \sim iut$ in the limit where $u \to 0$. The integral is finite. However, for large times ($\xi_0 t \gg 1$) this cutoff of the logarithmic divergence happens at lower and lower frequencies. This dependence can be approximately expressed as

$$F_2(t) \approx -g\int_{1/it}^{\xi_0} \frac{du}{u} \approx -g\ln(1 + it\xi_0) \tag{9.198}$$

The logarithm has the correct two limits of $i\xi_0 t$ at small time and as $\ln(it\xi_0)$ at large times. This approximation completes the evaluation of the renormalization factor:

$$\rho(t) = \exp[-itE_i - g\ln(1 + it\xi_0)] \tag{9.199}$$

Its Fourier transform determines the spectral function of the core hole:

$$A_h(\omega) = -2\,\mathrm{Im}\left[\int_{-\infty}^{\infty} dte^{i\omega t}G_h(t)\right] = 2\,\mathrm{Re}\left[\int_0^{\infty} dte^{i(\omega-\omega_T)t}\rho(t)\right] \tag{9.200}$$

The real part of the integral is the same as the integral plus its complex conjugate:

$$A_h = \int_0^{\infty} dte^{i(\omega-\omega_T)t}\rho(t) + \int_0^{\infty} dte^{-i(\omega-\omega_T)t}\rho^*(t) \tag{9.201}$$

The variable change $t \to -t$ in the second term, along with $\rho^*(-t) = \rho(t)$, brings us to the equivalent expression

$$A_h = \int_{-\infty}^{\infty} dte^{i(\omega-\omega_T)t}\rho(t) \tag{9.202}$$

$$= \int_{-\infty}^{\infty} dt\frac{e^{it\Omega\xi_0}}{(1 + it\xi_0)^g} \tag{9.203}$$

$$\Omega = \frac{\omega - \omega_T - E_i}{\xi_0} \tag{9.204}$$

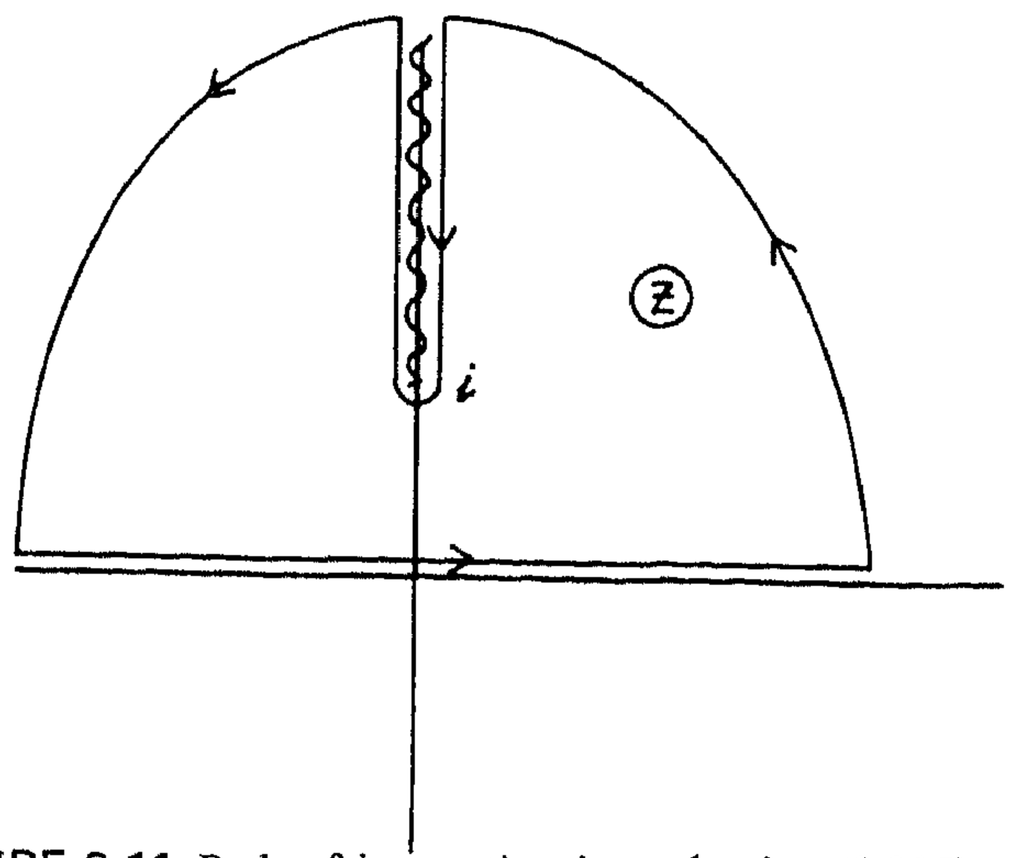


FIGURE 9.11 Path of integration in evaluating (9.205).

The integral is evaluated by a contour integration. Change the integration variable to $z = \xi_0 t$. There is a branch point at $z = i$. The branch cut is drawn vertically from this point, up the imaginary axis, as shown in Fig. 9.11:

$$A_h = \frac{1}{\xi_0}\int_{-\infty}^{\infty} dt \frac{e^{iz\Omega}}{(1+iz)^g} \tag{9.205}$$

For $\Omega < 0$, the integral $(-\infty, \infty)$ is closed by a contour in the lower half plane (LHP). This closed contour encloses no poles or branch cuts and so gives zero and shows that $A_h(\omega) = 0$ for $\Omega < 0$. The other situation is $\Omega > 0$, and here the contour of integration is closed in the UHP, as shown in Fig. 9.11. The contour encloses no poles, so the contribution on the real axis $(-\infty, \infty)$ is equal to that along the cut. For the integration along the branch cut, change variables to $z = i(1+y)$. The denominator in (9.205) becomes $(1+iz)^g = [1-(1+y)]^g = [-y]^g = y^g \exp(\pm i\pi g)$, where the choice of $(\pm)$ in the exponent depends on the side of the branch cut.

The integrals along the branch cut give

$$A_h(\omega) = \frac{\Theta(\Omega)}{\xi_0} i \int_0^{\infty} \frac{dy}{y^g} e^{-\Omega(1+y)}(e^{-i\pi g} - e^{i\pi g}) \tag{9.206}$$

$$= \Theta(\Omega)\Gamma(1-g)\frac{2\sin(\pi g)}{\xi_0}\frac{e^{-\Omega}}{\Omega^{1-g}} \tag{9.207}$$

The shape of this spectral function is shown in Fig. 9.12. It is zero for $\Omega < 0$ and nonzero for $\Omega > 0$. It diverges as a power law for $\Omega \sim 0^+$, with the exponent $(1-g)$. It is easy to cheek that the area under the spectral function is 2π:

$$1 = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A_h(\omega) = \frac{\sin(\pi g)}{\pi}\Gamma(1-g)\int_0^{\infty} d\Omega \frac{e^{-\Omega}}{\Omega^{1-g}} \tag{9.208}$$

$$1 = \frac{\sin(\pi g)}{\pi}\Gamma(1-g)\Gamma(g) \tag{9.209}$$

The last equation is an identity among gamma functions.

The response of the electron gas has a dramatic effect upon the spectral function of the hole. The spectral function has a power law divergence Ω^{1-g} near the threshold energy. In t absence of interactions $(V = 0)$, the spectral function is a delta function $A_h(\omega)$ $2\pi\delta(\omega - \omega_T)$. As illustrated schematically in Fig. 9.13(a), the influence of interactions is

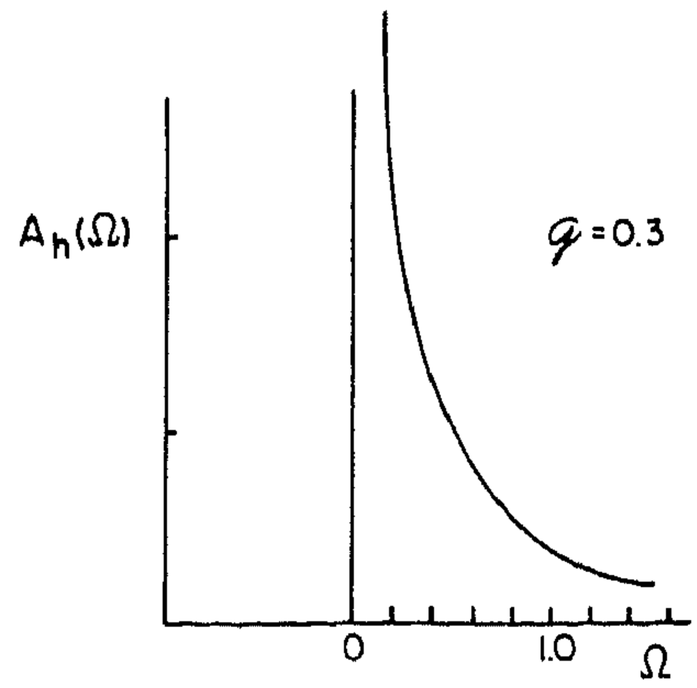


FIGURE 9.12 Spectral function of the hole Green's function. The orthogonality catastrophe causes the power law distribution and eliminates any delta function character.

change the delta function into a power law divergence. All quasiparticle behavior, in the form of a remnant delta function, has vanished.

In X-ray absorption, the unperturbed spectrum for $V = 0$ is a step $\Theta(\omega - \omega_T)$ rather than a delta function, which is illustrated in Fig. 9.13(b). How does this step become modified by the potential V in the orthogonality catastrophe? The answer is that the threshold becomes a power law Ω^g which goes to zero as $\Omega \to 0$. The threshold behavior may be derived in the following way. The Fourier transform of a step function $\theta(\omega - \omega_T)$ is t^{-1}:

$$\Theta(\omega - \omega_T) = \int_{-\infty}^{\infty} \frac{dt}{2\pi i} \frac{e^{it(\omega - \omega_T)}}{t - i\delta} \tag{9.210}$$

The orthogonality catastrophe modified this spectrum by the inclusion of the term $\rho(t)$ in the integrand. Call the resulting X-ray spectra $X(\omega)$:

$$X(\omega) = \int_{-\infty}^{\infty} \frac{dt}{2\pi i} \frac{1}{t - i\delta} \frac{e^{it\Omega\xi_0}}{(1 + it\xi_0)^g} \tag{9.211}$$

The easiest way to evaluate this integral is to note that its derivative is just the hole spectral function

$$\frac{dX(\omega)}{d\omega} = \frac{1}{2\pi} A_h(\omega) \tag{9.212}$$

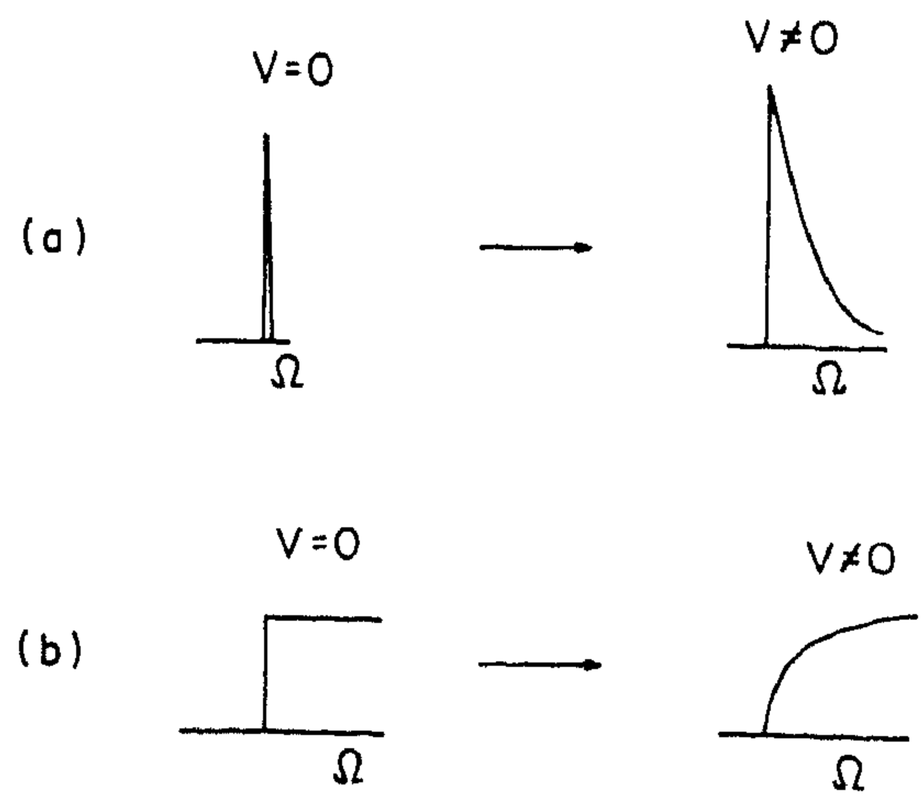


FIGURE 9.13 The effects of the orthogonality catastrophe upon spectral distributions. (a) The delta function of a spectral function is changed to a power law divergence. (b) The step function in absorption is changed to a converging threshold.

This equation has the solution

$$X(\omega) = \int_{-\infty}^{\omega} \frac{d\omega'}{2\pi} A_h(\omega') = \frac{\Theta(\Omega)}{\Gamma(g)} \int_0^{\Omega} d\Omega' \frac{e^{-\Omega'}}{(\Omega')^{1-g}} \tag{9.213}$$

The behavior near threshold may be deduced by examining the integral for very small values of Ω. Then the slowly varying term $\exp(-\Omega')$ can be ignored, and the integral gives

$$X(\omega) = \frac{\Theta(\Omega)}{\Gamma(g)} \int_0^{\Omega} \frac{d\Omega'}{(\Omega')^{1-g}} = \frac{\Omega^g \Theta(\Omega)}{\Gamma(1-g)} \tag{9.214}$$

Equation (9.214) shows that the absorption behaves as Ω^g near threshold. The threshold has a converging behavior, with the absorption starting at zero when $\Omega = 0$. It is sketched in Fig. 9.13(b).

The absorption is zero at the threshold frequency $\Omega = 0^+$ because of the orthogonality catastrophe. The transition is not allowed because of orthogonality and has a zero probability. The absorption is nonzero for $\Omega > 0$ because of the creation of a number of electron–hole pairs in the system. This serves as a symmetry-breaking process, which makes the transition allowed.

The role of electron–hole pairs in the X-ray transition was elucidated by Schotte and Schotte (1969). The interaction term between the core hole and the electron gas can be written in terms of the electron density operator

$$\rho(\mathbf{q}) = \sum_{\mathbf{k},s} C^{\dagger}_{\mathbf{k}+\mathbf{q},s} C_{\mathbf{k}s} \tag{9.215}$$

$$\frac{1}{\nu} \sum_{\mathbf{k}\mathbf{k}'ss'} C^{\dagger}_{\mathbf{k}\sigma} C_{\mathbf{k}'s} d^{\dagger}_{s'} d_{s'} = \frac{1}{\nu} \sum_{\mathbf{q},s'} V(\mathbf{q}) \rho(\mathbf{q}) d^{\dagger}_{s'} d_{s'} \tag{9.216}$$

The density operator can be represented by the boson operators for the electron–hole excitations of the electron gas $\rho(\mathbf{q}) \to (b_{\mathbf{q}} + b^{\dagger}_{-\mathbf{q}})$. The Hamiltonian between the core hole and the electron gas is transformed into a problem of a localized level and a boson system. This model Hamiltonian can be solved exactly as the independent boson model of Sec. 4.3. The solution is entirely analogous to the phonon core hole system. The response of the boson system to the sudden appearance of the core hole is described by including the factor of $\exp[-\phi(t)]$ in the time integrals The factor $\phi(t)$ for phonons is presented in (9.144) and is rewritten as

$$\begin{aligned} \phi(t) &= i\Sigma t + \int_0^{\omega_D} \frac{du}{u^2} R_{ph}(u) \{[n_B(u) + 1](1 - e^{-iut}) + n_B(u)(1 - e^{iut})\} \\ R_{ph}(u) &= \sum_{\mathbf{q}\lambda} |M_{\lambda}(\mathbf{q})|^2 \delta[u - \hbar\omega_{\lambda}(\mathbf{q})] \end{aligned} \tag{9.217}$$

The self-energy term $i\Sigma t$ is put first, while the other term is the transient behavior. The approximate response of the electron system to the core hole was given earlier as $F_2(t)$ in

(9.192). It has exactly the same form as the preceding, except the phonon system is at nonzero temperature. The electron–electron response at nonzero temperature is

$$\phi(t) = -F_2(t) = \int_0^\infty \frac{du}{u^2} R_e(u)\{[n_B(u)+1](1-e^{-iut}) + n_B(u)(1-e^{iut})\}$$

$$R_e(u) = -\frac{1}{\pi}\int \frac{d^3q}{(2\pi)^3} V(q)^2 \operatorname{Im}[P^{(1)}(\mathbf{q}, u)] \tag{9.218}$$

The quantity $R_{ph}(u)$ for the phonon system is the phonon density of states, which is weighted by the matrix element for the process. Similarly, the factor $R_e(u)$ serves the same role for electron–electron interactions. It is interpreted as the effective coupling to the electron hole pairs. The equations are exactly alike for the response of the core hole phonon and core hole pairs, and enter into the spectral function for absorption in the same fashion as factors $\exp(-\phi(t))$. This similarity is reasonable, since both the phonons and electron–hole pairs are boson systems which respond in the same way to the appearance of the core hole.

The only difference between these two boson systems is their respective coupling constants $R_{ph}(u)$ and $R_e(u)$. This difference is big, since they behave quite differently at small values of u. The electron–electron interactions were shown to have a linear relation $R_e \propto u$ which leads to power law behavior in $A(\omega)$. The phonon system is governed by the acoustic phonons at small values of u and usually behaves as $R_{ph} \propto u^2$ in metals. Consequently, the exponential factor for the phonon system does not diverge as $\sim \ln(it\xi_0)$ at large $\xi_0 t$ but goes to a constant except for the self-energy term. The spectral function $A(\omega)$ from the phonon part alone is a dull function, which was represented as a Gaussian. The core hole phonon and core pair couplings cause quite different behaviors in the time response of the system and affect the spectral shape differently. The difference arises from the dissimilar nature of their coupling function $R_{ph}(u)$ and $R_e(u)$.

The electron–electron response is an example of a type of phenomenon known as an *infrared divergence*. The divergence occurs in the limit where $u \to 0$, which is the infrared part of the spectrum. The quantity $R_j(u)/u^2$ is the probability (at $T = 0$) of emitting an excitation of energy u during the sudden switching on of the core bole potential. For example, the number Nu of excitations created is

$$\mathrm{Nu} = \int_0^\infty \frac{du}{u^2} R_e(u) \sim g\int_0^{\xi_0} \frac{du}{u} = -g\ln(0) \tag{9.219}$$

The average number Nu of bosons around the particle was discussed earlier for polarons in Sec. 4.1. It is infinity for the electron–hole excitations, which shows that an infinite number are created. This infinity does not imply a physical disaster, since all measurable quantities are nonzero. For example, the energy which is released by the creation of these electron–hole pairs is

$$\Delta E = \int_0^\infty \frac{du}{u^2} R_e(u) u \sim g\int_0^{\xi_0} du = g\xi_0 \tag{9.220}$$

which is nonzero. However, a divergence in the number Nu of excitations is usually an indication that the spectral function will have an unusual shape, since quasiparticle behavior is eliminated. A further discussion is given by Hopfield (1969) and Mahan (1974). The phonon system does not have an infrared divergence in the X-ray problem in metals, although it can in semiconductors where the piezoelectric interaction is not screened.

The exponential series for $F(t) = \sum_i F_i(t)$ in (9.183) has been approximated by the term $F_2(t)$ plus $iE_i t$. Further terms in the series also contribute to the orthogonality catastrophe and should be summed in order to obtain an accurate description of the phenomenon. The exact method of summing all of these terms has been given by Mahan (1982), and more elegantly by Ohtaka and Tanabe (1983). They provide an exact expression for $R_e(u)$.

Several other types of results have been obtained. The most important is by Nozieres and deDominicis (1969), who show that the exact coefficient of the term which diverges like $\ln(it\xi_0)$ is

$$\lim_{t\to 0} F(t) = -itE_i + \alpha \ln(it\xi_0) \tag{9.221}$$

$$\alpha = \sum_{m_s, m_l, l} \left[\frac{\delta_l(k_F)}{\pi}\right]^2 \tag{9.222}$$

The same phase shifts $\delta_l(k_F)$ were mentioned earlier and arise from the scattering of the conduction electron. with wave vector k_F from the core hole. The core hole is in the center of a spherical coordinate system, and the conduction electron states are described by the quantum numbers of m_s, l, m_l and k. The phase shifts are usually taken to be independent of m_s and m_l, in which case the coefficient becomes

$$\alpha = 2\sum_l (2l+1)\left[\frac{\delta_l(k_F)}{\pi}\right]^2 \tag{9.223}$$

The coefficient α should be used in describing the asymptotic behavior of the orthogonality catastrophe very near the threshold.

The orthogonality catastrophe is an important feature of many-electron physics. The process of switching on a potential occurs in many circumstances, and the physics can be applied to a variety of phenomena. For example, the Kondo effect has this feature, which plays an important role in its theory. Whenever the potential is switched on, the system responds by making electron–hole pairs, which affects the spectral response of all correlation functions which are being evaluated.

9.3.4. MND Theory

The absorption of X-rays in a metal causes several different types of excitations and phenomena. Two have been discussed which cause power law behavior at the absorption edges: excitons and the orthogonality catastrophe. An important contribution was made by Nozieres and deDominicis (1969), who solved the two processes together. Their solution is asymptotically exact near threshold $\tilde{\omega}_T = \omega_T + E_i$. It expresses the power law exponents as functions of the angular momentum l of the conduction electron in its final state. The many-body correction to formula (9.142) with two angular momentum channels $A_{l\pm 1}$ is

$$A(\omega) = \theta(\omega - \tilde{\omega}_T)\left[A_{l+1}\left(\frac{\xi_0}{\omega - \tilde{\omega}_T}\right)^{\alpha_{l+1}} + A_{l-1}\left(\frac{\xi_0}{\omega - \tilde{\omega}_T}\right)^{\alpha_{l-1}}\right]$$

$$\alpha_l = \frac{2\delta_l(k_F)}{\pi} - \alpha \tag{9.224}$$

$$\alpha = 2\sum_l (2l+1)\left[\frac{\delta_l(k_F)}{\pi}\right]^2 \tag{9.225}$$

The power law exponents α_l are the sum of two terms. The first is $2\delta_l/\pi$, which is from the exciton phenomenon; this term is usually positive in free-electron metals and tends to make the edges diverge and become singular. The other factor α comes from the orthogonality catastrophe and tends to make the edges converging at threshold. The final value of α_l for an absorption edge depends on the difference of these two quantities and may be either positive or negative.

There have been many calculations of these phase shifts and exponents for comparison with those deduced from experiments. Typical results for sodium are shown in Table 9.2, as calculated by Minnhagen (1977). A recent calculation is by Ohmura and Ogiwara (1994). Minnhagen used a pseudopotential, which was screened by a Singwi–Sjölander dielectric function. The phase shifts should always obey the Friedel sum rule in Sec. 4.1,

$$1 = \frac{2}{\pi}\sum_l (2l+1)\delta_l(k_F) \tag{9.226}$$

which is just a statement that the screening around the core hole must be one unit of charge. The exponents are typical for simple metals. The X-ray absorption from a p shell, where the initial angular momentum is $l = 1$, has the final values for the conduction electron of $l = 0$ and $l = 2$. The asymptotic limit of the Mahan–Nozieres–deDominicis (MND) theory for $\omega > \tilde{\omega}_T$ is

$$A(\omega) = \theta(\omega - \tilde{\omega}_T)\left[A_0\left(\frac{\xi_0}{\omega - \tilde{\omega}_T}\right)^{\alpha_0} + A_2\left(\frac{\xi_0}{\omega - \tilde{\omega}_T}\right)^{\alpha_2}\right]$$

The first term is the s-wave channel. The exponent α_0 is positive, so this term diverges at threshold. The second is the d-wave channel where $\alpha_2 < 0$ and the term goes to zero at threshold. The summation of these two terms diverges at threshold. The theory predicts that absorption edges are divergent for electrons from p shells, which is generally observed in free-electron metals such as Na, Mg, Al, and K. On the other hand, if the initial core state is an s shell with $l = 0$, the final state of the conduction electron has $l = 1$, $\alpha_1 \approx -0.05$ for sodium. This value is so small that it predicts there is no measurable effect from the many-electron response. This prediction is also in accord with experimental observations. The value of α_1 is small because of the partial cancellation of $2\delta_1/\pi = 0.14$ and $\alpha = 0.19$. It seems to be a general feature of free-electron metals that the transitions from core hole s states shows no additional peaking or rounding of the threshold, in agreement with the predictions that α_1 is very small. MND theory seems to qualitatively explain the behavior of the edges of simple metals.

Figure 9.14 shows the experimental results of Callcott *et al.* (1978) for the absorption edge of the $L_{2,3}$ shell of metallic sodium. The p-shell is split by the spin–orbit interaction into the $p_{3/2}$ and $p_{1/2}$ components, which comprise the L_3 and L_2 shells. In sodium, this splitting is

TABLE 9.2 Phase shifts and exponents for sodium. $\alpha = 0.19$. From Minnhagen (1977) (used with permission)

l	δ_l	α_l
0	0.90	0.38
1	0.22	−0.05
2	0.01	−0.19

0.17 eV and accounts for the double edge shown in absorption. Each of these two edges should have a power law singularity. Figure 9.14 also shows the deconvolution of the sodium spectrum into the part due to the one-particle spectrum and broadening processes such as the Lorentzian width from Auger and the Gaussian width from phonons. The absorption edges show definite spikes which are from the edge singularities. The spikes are much smaller in emission, although that is partly from the increased phonon width in emission due to the incomplete relaxation phenomenon, as mentioned earlier. The value of the edge singularity exponent deduced by Callcott *et al.* to fit their absorption data is $\alpha_0 = 0.25$, which is smaller than the $\alpha_0 = 0.38$ predicted from the phase-shift calculations. The theoretical values of α_0 have been computed by many different workers, and values similar to those shown in Table 9.2 are obtained by many different groups, so the theoretical values seem reliable.

There have been several explanations for this lack of detailed agreement between the measured and theoretical exponents: (1) omission of frequency dependent screening in the theory; (2) exchange type of effects between the core hole spin and electron spin, as discussed by Girvin and Hopfield (1976); (3) scattering of the core hole among its different orbital states in the L shell; and (4) the asymptotic solution (9.217) is not valid very far from threshold. This subject is still an active area of research. A summary of the present status of this field seems to be that the edge singularities are observed in some thresholds of simple metals, and the refinements of the theory are still under investigation.

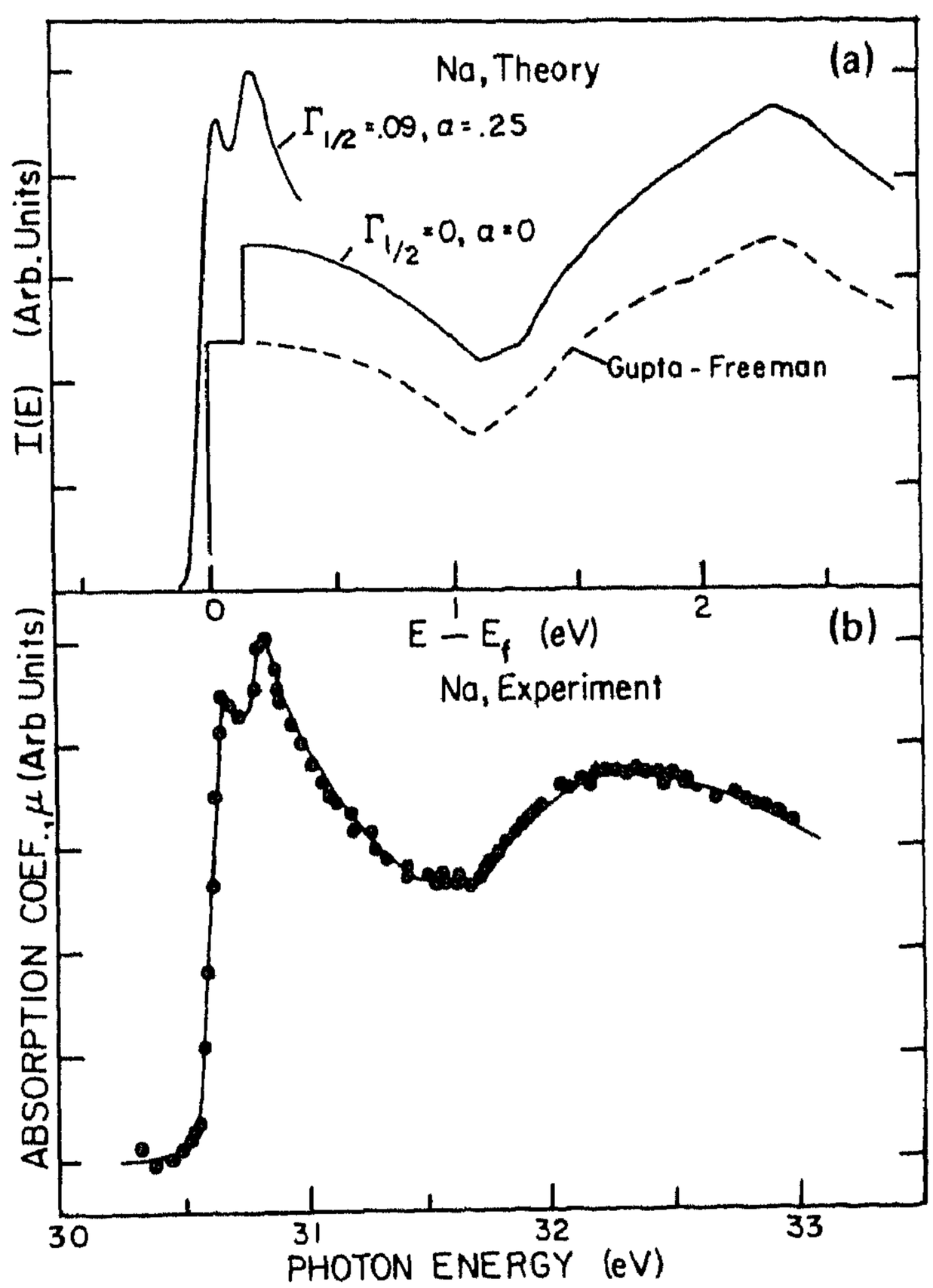


FIGURE 9.14 The X-ray absorption edge of the $L_{2,3}$ shell of metallic sodium. The data is shown in part (b). Part (a) shows how the band structure prediction of Gupta and Freeman (1976) is convoluted with the edge singularity ($\alpha = 0.25$) and broadening functions to fit the spectra. *Source*: Callcott *et al.* (1978) (used with permission).

9.3.5. XPS Spectra

X-ray photoelectron spectroscopy (XPS) measures the line shapes of photo-electrons excited from the core levels of atoms. A photon, which is usually in the kilovolt energy range, excites a core electron to very high kinetic energy. Since the binding energies E_B of the L-shell electrons in Na, Mg, and Al are only 50–100 eV, the final kinetic energy of the election is $E_f = \hbar\omega - E_B$. In a one-electron model, all the photoelectrons excited by the same frequency ω would have exactly the energy E_f, and a measurement of the distribution of kinetic energies of the photoelectrons would show a delta function at this value. The actual kinetic energies of photoelectrons, measured from simple metals, show the main peak at energy $E \sim E_f$ plus satellite peaks which correspond to the emission of bulk and surface plasmons (Pardee *et al.*, 1975). The main peak at $E \sim E_f$ has a nonzero width, which must be partly caused by the usual mechanisms of (1) Auger lifetime of the core hole and (2) phonon broadening due to the coupling between the core hole and the phonons.

Doniach and Sunjic (1970) predicted that the orthogonality catastrophe should make the shape of the main line, at $E \sim E_f$, asymmetric with an asymmetry parameter which is a direct measurement of the index α. Their argument is that this experiment directly measures the spectral function of the core hole, since the outgoing electron has too much energy to be affected by the interaction with phonons or electron–hole pairs. In the calculation of the Kubo formula for the absorption rate, the conduction electron operators $[C^\dagger_{\mathbf{p}\sigma}, C_{\mathbf{p}\sigma}]$ can be removed from the correlation function on the grounds that the electrons leave the core hole too rapidly to be influenced by exciton or other final state processes. The correlation function is factored:

$$A(\omega) = \frac{1}{\omega}\int_{-\infty}^{\infty} dt e^{i\omega t}\langle \mathscr{J}(t)\mathscr{J}(0)\rangle \tag{9.227}$$

$$= \frac{1}{\omega}\int_{-\infty}^{\infty} dt e^{i\omega t} \sum_{\mathbf{k}\mathbf{k}'ss'} w(\mathbf{k})^* w(\mathbf{k}')\langle C_{\mathbf{k},s}(t)C^\dagger_{\mathbf{k}'s'}(0)\rangle\langle d_s(t)d^\dagger_{s'}(0)\rangle$$

$$= \frac{2}{\omega}\int_{-\infty}^{\infty} dt \sum_{\mathbf{k},s'} e^{it(\omega-\xi_{\mathbf{k}})} w(\mathbf{k})^2\langle d_{s'}(t)d^\dagger_{s'}(0)\rangle \tag{9.228}$$

The Kubo formula for the absorption becomes proportional to the spectral function $A_h(\omega - \xi_{\mathbf{k}})$ for the core hole. In this experiment, the photon energy ω is usually fixed, and measurement is made of the distribution $P(E)$ of final kinetic energies $E \equiv \xi_k - E_W$ of the conduction electron, where E_W is the work function of the metal. XPS spectra can now be measured with millivolt energy resolution, although the electron energy E is often over a kilovolt. The matrix elements $w(\mathbf{k})$ are a constant over the narrow energy range of $\xi_{\mathbf{k}}$ which are measured for each XPS line. The probability $P(E)$ is just the spectral function of the core hole $A_h(\omega - E - E_W)$

$$P(E) = w^2 f(E - E_f) \tag{9.229}$$

$$f(E) = \int_{-\infty}^{\infty} dt \exp[-itE - \tfrac{1}{2}\Gamma|t| - \gamma^2 t^2 - \alpha \ln(it\xi_0)] \tag{9.230}$$

where $E_f = \hbar\omega - E_W$. It has the power law singularity in (9.207) and is illustrated in Fig. 9.12, which is then broadened by the phonons and the Auger decay. The line shape is predicted to be asymmetric because of the orthogonality catastrophe, and the degree of asymmetry is just given by the index α. The formula used for interpreting the experimental

line-shape function $f(E)$ contains the Lorentzian width from Auger processes, the Gaussian width from phonons, and the asymmetry from the electron–hole pair creation.

An early experimental result is shown in Fig. 9.15 from data of Citrin (1973) for metallic sodium. The dashed line indicates the excess line shape on the low-energy side. Many more spectra have been published by Citrin *et al.* (1977). One feature of the experimental results is that the same index α fits the asymmetry for different core holes in the same atom: the same α is found for the different shells K, L_2, L_3, etc. The ion with a core hole looks the same, to electron–hole pairs which are outside the ion, regardless of the particular state of the core hole.

The XPS line shape is asymmetric, because of the emission of electron–hole pairs during the creation of the core hole. The photon energy is fixed at $\hbar\omega$, and any energy used to make the pairs must be subtracted from that carried away by. the conduction electron. The conduction electron has a high probability of having energy less than E_f. This explanation accounts for the sign of the asymmetry and explains why the line tails to the left in Fig. 9.15. These asymmetric line shapes had been observed for many years, as reviewed by Parratt (1959). He speculated then that the "excess width on the low energy side is essentially

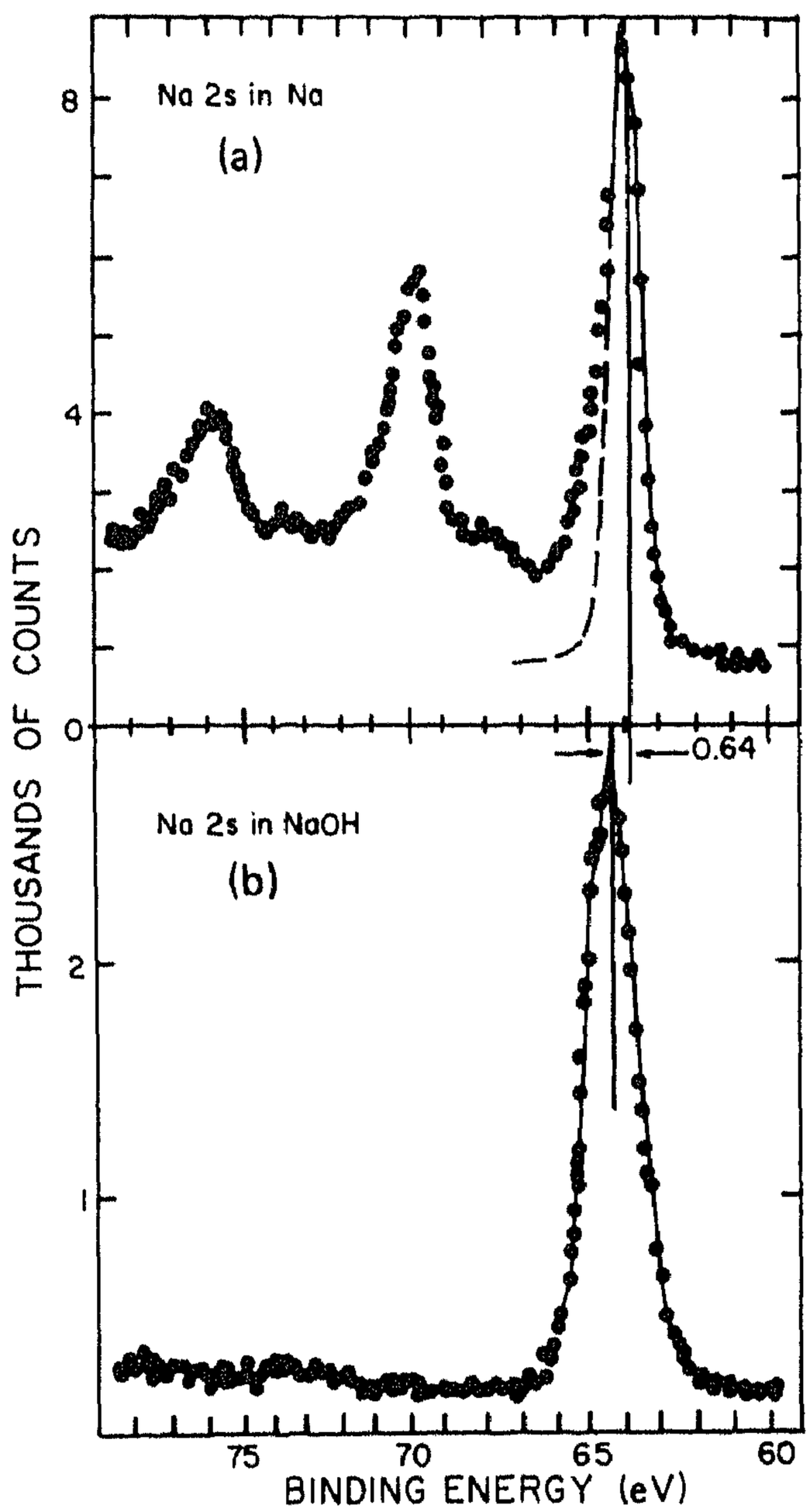


FIGURE 9.15 X-ray photoelectron spectra from sodium 2*s* electrons in (a) metal and (b) hydroxide. The metal spectra is asymmetric, with additional contributions on the low-energy side due to pair production. These are lacking in the insulator. Peaks to the left are plasmons in the metal. *Source:* Citrin (1973) (used with permission).

attributable to transitions between excitation states of the valence-electron-configuration type." The quantitative theory of Doniach and Sunjic has verified this hypothesis in great detail and provides the most direct method of measuring the orthogonality catastrophe.

PROBLEMS

1. Derive the nonzero temperature form of the free-polaron absorption (9.51) assuming Maxwell–Boltzmann statistics. To order α, an exact result can be obtained in terms of Bessel functions $K_1(\frac{1}{2}\beta|\omega \pm \omega_0|)$.

2. Consider the force–force correlation function for scattering from an impurity. Discuss whether the multiple. scattering from an impurity can be represented by a T matrix or similar series.

3. Derive the Wilson–Butcher formula (9.64) from the golden rule. Find the matrix element $\langle f|\hat{\varepsilon} \cdot \mathbf{p}|i\rangle$ by writing both initial $|i\rangle$ and final $|f\rangle$ Bloch states to first order in the potential V_G.

4. Consider a Hamiltonian which is the homogeneous electron gas plus the crystal potential $\sum V_{\mathbf{G}}\rho(\mathbf{G})$. Discuss the summation of terms which occurs in higher order when evaluating the force–force correlation function.

5. Where do the final state interactions for the Coulomb scattering of the electron and hole appear in formula (9.64) for interband transitions?

6. Show that the rate of photon emission in an insulator with a nonequilibrium distribution of electrons and holes is governed by the matrix element $|w_f(\mathbf{k})|^2$ as in (9.132). Assume the initial state of the system at zero temperature is $|k\rangle = d^\dagger a_k^\dagger|0\rangle$, and use arguments analogous to those following (9.139).

7. Write out the correlation function $\pi^{(1)}$ in (9.105), with one vertex diagram, for the frequency-dependent screening function (7.171). Do the Matsubara summations.

8. Show there is an infrared divergence in the X-ray response resulting from the piezoelectric electron–phonon interaction in insulators (Sec. 1.3) when used in the response function (9.217).

9. Consider the X-ray edge problem with an interaction term H_{sd} in Sec. 1.4.2 between the conduction electrons and the core hole. What is the contribution of this term to the orthogonality index α from the cumulant $F_2(t)$ (Girvin and Hopfield, 1976)?

10. Show that the imaginary part of the retarded correlation function (8.34) has the following sum rule relating to the average kinetic energy:

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} n_B(-\omega)\mathrm{Im}[\pi(\omega)] = \frac{e^2 n_0}{m} E_{KE} \tag{9.231}$$

Chapter 10

Superconductivity

The theory of superconductivity was formulated by Bardeen, Cooper, and Schrieffer (1957) and is called the BCS theory. It very successfully describes the superconducting properties of *weak* superconductors, such as aluminum, which are weak because of the small strength of the electron–phonon interaction. Further refinements of the theory have led to the strong coupling theory of Eliashberg (1960) which describes well the properties of superconductors such as lead. The distinction between aluminum and lead is roughly determined by the value of the electron–phonon mass enhancement factor λ, as shown by McMillan (1968). The BCS theory is discussed first. It must rank as one of the great successes of many-body formalism, since the theory provides detailed agreement with experiments. This agreement is a refreshing change from most comparisons between many-body theory and experiment, where the results often depend upon vertex corrections, correlations, and computer simulations. The beauty of BCS is that it is, mathematically, a simple theory which is exceedingly accurate. The reason it works is that the basic coupling forces are weak, and mean field works well.

The basic idea of BCS theory is that the electrons in the metal form bound pairs. Not all electrons do this, but only those within a Debye energy of the Fermi surface. The bound states of the electron pairs are not described by simple orbitals such as used for the hydrogen atom or positronium. The pair state, and the entire ground state of the superconductor, requires a many-body description.

The Debye energy enters into the ground state description because the attractive forces between electrons, which are responsible for the pair binding, are due to the electron–phonon interaction. Fröhlich (1950) was the first to realize that electrons could interact by exchanging phonons and that this interaction could be attractive. He was the first to suggest that superconductivity was caused by the electron–phonon interaction. The phonon dependence would explain the experimental observation that the transition temperature T_c is a function of the ion mass for different isotopes of the same metal. This isotope effect was discovered for the metal Hg by Maxwell (1950) and Reynolds *et al.* (1950), where the dependence was proportional to $\Delta T_c/T_c \approx -\delta M/(2M)$. The BCS theory explains this in detail, since it shows that the transition temperature is proportional to the Debye frequency $k_B T_c \approx \omega_D \propto M^{-1/2}$. The isotope effect verified the Fröhlich hypothesis that the electron–phonon interaction caused superconductivity.

Another piece in the theoretical puzzle was supplied by Schafroth (1955), who showed that a charged boson gas, when undergoing a Bose–Einstein condensation, would exhibit many of the superconducting properties known at that time such as the Meissner effect. His

theory does not explain other features, such as the energy gap in the excitation spectrum. Schafroth speculated that in the superconductor the "bosons are resonant two electron states." The BCS theory uses a similar mechanism, since the paired electrons behave, in some respects, as bosons. The BCS theory was not conceived in a vacuum but among many related ideas, each of which contained some element of truth.

BCS was the first theory to explain superconductivity in metals and also made a number of remarkable predictions. The foremost was that an energy gap existed in the excitation spectrum of the superconductor. The actual observation of this energy gap by electron tunneling (Giaever, 1960) provided a dramatic verification of the theory, although BCS earlier argued that the thermodynamic data supported the existence of a gap. Many different experiments in weak superconductors have shown that the original version of the theory is correct in its many details. An extensive comparison between theory and experiment is provided in the two volumes of *Superconductivity* edited by R.D. Parks (1969) and in the books by Rickayzen (1965) and Schrieffer (1964).

In 1987 superconductivity was discovered in materials with planes of copper oxide, which is now the field of high-temperature superconductivity. The highest value of the transition temperature at ambient pressure is $T_c = 133\,\mathrm{K}$ in the Hg–Ba–Ca–Cu–O system (Schilling *et al.*, 1993). Although much is known about these materials, the basic mechanism causing superconductivity at high temperatures is still unexplained. Even the symmetry of the energy gap is uncertain. Most experiments support the d-wave picture, but some still support s-wave symmetry. The electron seemed to be paired in spin singlet states, with a very short coherence length. It is also interesting that experiments which were important in understanding elemental superconductors, such as electron tunneling or infrared absorption, proved to be difficult to interpret for the cuprates.

10.1. COOPER INSTABILITY

The first inkling of the BCS theory was a letter by Cooper (1956), who pointed out that the ground state of a normal metal was unstable at zero temperature. A normal metal is defined as one which is neither superconducting nor magnetic. The instability is an indication that the metal prefers to be in another state, in this case the superconducting one. The demonstration of an instability does not provide a description of the superconducting state, but it did suggest that the instability was caused by the scattering between pairs of electrons, where the scattering potential is the exchange of phonons.

Two electrons can scatter as shown in Fig. 10.1. The screened interaction between two electrons was derived in Sec. 6.4:

$$V_s(\mathbf{q}, \omega) = \frac{v_q}{\varepsilon(\mathbf{q}, \omega)} + \frac{2\omega\lambda\, M_{\mathbf{q}}^2}{\varepsilon(\mathbf{q})^2[\omega^2 - \omega_\lambda(\mathbf{q})^2]} \tag{10.1}$$

It has two terms. The first is the screened Coulomb interaction. The theory of superconductivity is applied at low temperatures, where the energy exchanged between particles, while scattering, is also low. Requirements of crystal stability require that this interaction be repulsive at zero frequency (Allen *et al.*, 1988). There have been numerous attempts to bind two electrons using just the first term which has only an electron–electron interaction. Then the plasmon is the boson which is exchanged between electrons. Careful calculations,

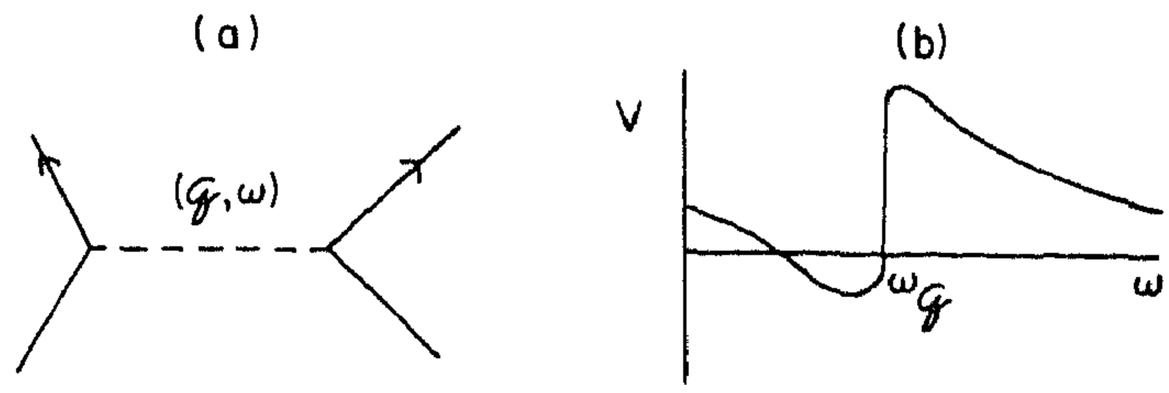


FIGURE 10.1 The pairing force in a superconductor is due to the exchange of excitations, such as phonons, between pairs of electrons. This interaction is frequency dependent and has attractive regions for nonzero frequencies.

including local field corrections, show that the binding force from plasmon exchange is negligibly small.

The second term in (10.1) is the screened electron–phonon interaction. It is, on the average, weaker than the repulsive Coulomb interaction. However, for frequencies near to the Debye ($\omega < \omega_D$) the energy denominator becomes small and negative, which causes a relatively large interaction over this narrow range of frequency. The frequency dependence of $V_s(\mathbf{q}, \omega)$ is illustrated schematically in Fig. 10.1(b). It may be possible for two electrons to bind if they can construct a relative wave function which selectively uses the frequency region which is attractive. Not all metals are superconductivity; the alkali metals are not. They have a weak electron–phonon interaction. So the attractive interaction needs some minimum strength in order to cause superconductivity.

There are two ways to proceed with the theory. One is to use the full frequency-dependent interaction, which sums over all phonon wave vectors and modes. This realistic calculation is done in a later section. The other possibility is to replace (10.1) with a model interaction of the form

$$V_s(\mathbf{q}, \omega) = \begin{cases} -V_0 & \text{for } |\xi_q| < \omega_D \\ 0 & \text{for } |\xi_q| > \omega_D \end{cases} \tag{10.2}$$

This potential is constant and attractive ($V_0 > 0$) up to a cutoff energy which is of the order of the Debye energy ω_D of the solid. This second form of the interaction permits a much simpler discussion of the theory and allows the physics to be introduced more easily. The initial discussion of superconductivity uses the model potential in (10.2), which follows the historical pathway of Cooper and BCS. Later, when the physics is better understood by the reader, the theory is redone from the beginning using the realistic potential (10.1). The latter is called *strong-coupling theory.*

Cooper's model of a normal metal at low temperature was a free-electron system. In the limit of zero temperature, the Fermi surface has a sharp step in energy. The electrons are allowed to have a weak attractive interaction as given in (10.2). Consider the mutual scattering of two electrons. Assume they initially have states of equal and opposite momentum $\mathbf{k}$ and $-\mathbf{k}$, or zero center of mass. Why this is necessary for the Cooper instability is shown later. It is also assumed the particles have opposite spin states $\uparrow$ and $\downarrow$, so that exchange scattering does not occur. The interaction potential does not flip the electron spin and the spin states are preserved in the scattering process.

Figure 10.2 shows a double scattering event between two electron lines which are moving in the same direction in time. This process is the scattering in the second Born approximation, where the first Born approximation is shown in Fig. 10.1(a). Each dashed line represents an interaction of the type shown in (10.2). If two electrons initially start in opposite momentum states $\mathbf{k}$ and $-\mathbf{k}$, then a momentum transfer $\mathbf{q}$ leaves them still in opposite

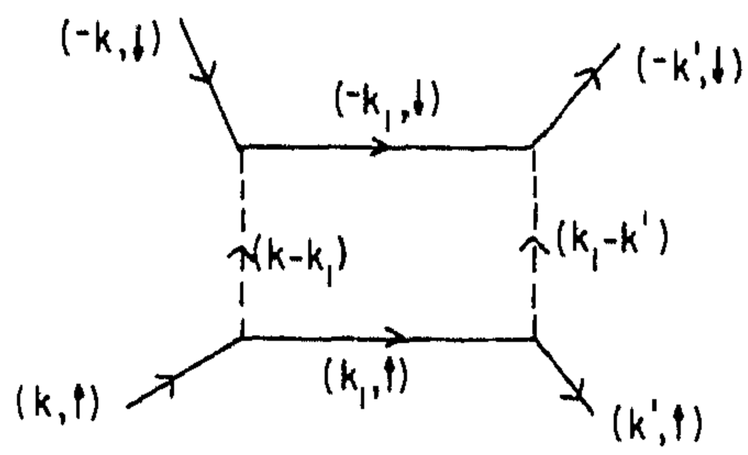


FIGURE 10.2

momentum states $\mathbf{k}+\mathbf{q}$ and $-(\mathbf{k}+\mathbf{q})$. This process is shown in the figure as the pairs $(\mathbf{k}_1, -\mathbf{k}_1)$ in the intermediate state and $(\mathbf{k}', -\mathbf{k}')$ in the final state of the double scattering. The effective scattering in the first and second Born approximation is

$$V_{\text{eff}}(\mathbf{k}-\mathbf{k}') = V(\mathbf{k}-\mathbf{k}') + \int \frac{d^3k_1}{(2\pi)^3} \frac{V(\mathbf{k}-\mathbf{k}_1)V(\mathbf{k}_1-\mathbf{k}')}{2\xi_{\mathbf{k}} - 2\xi_{\mathbf{k}_1}} \times \{[1-n_F(\xi_{\mathbf{k}_1})][1-n_F(\xi_{\mathbf{k}_1})] - n_F(\xi_{\mathbf{k}_1})^2\} \tag{10.3}$$

The second term on the right is the contribution of Fig. 10.2. The energy denominator contains the initial state energy $\xi_{\mathbf{k}} + \xi_{-\mathbf{k}} = 2\xi_{\mathbf{k}}$ minus the intermediate state energy $2\xi_{\mathbf{k}_1}$. The numerator contains the important factors $[1-n_F(\xi_{\mathbf{k}_1})]^2 - n_F(\xi_{\mathbf{k}_1})^2 = 1 - 2n_F(\xi_{\mathbf{k}_1})$ which occur because the two particles, which are being scattered, can only go into the states $(\mathbf{k}_1, \uparrow)$ and $(-\mathbf{k}_1, \downarrow)$ if they are not already occupied by an electron. This explains the factors $[1-n_F(\xi_{\mathbf{k}_1})]^2$. The other term $n_F(\xi_{\mathbf{k}_1})^2$ represents the scattering back into this state, $(\mathbf{k}_1, \uparrow) \to (\mathbf{k}, \uparrow)$, etc., since the result depends upon the net scattering. What is left are the remaining factors of $1 - 2n_F(\xi_{\mathbf{k}_1})$. These occupation factors of $n_F(\xi_{\mathbf{k}_1})$ play a crucial role in the theory and are the cause of the instability.

The integral in (10.3) may be evaluated. The key is that the interaction acts only over a small energy interval near the Fermi energy. Over this interval, of a Debye energy, the electron density of states in most metals is nearly constant. Change integration variables to $\int d^3k_1/(2\pi)^3 = \int d\xi_1 N(\xi_1)$ and treat $N(\xi \approx 0) \equiv N_F$ as a constant. At zero temperature the result is $(\xi_{\mathbf{k}_1} = \xi_1, \xi_{\mathbf{k}} = \xi)$

$$V_{\text{eff}}(\mathbf{k}-\mathbf{k}') = V(\mathbf{k}-\mathbf{k}') + N_F V_0^2 \int_{-\omega_D}^{\omega_D} d\xi_1 \frac{\frac{1}{2} - n_F(\xi_1)}{\xi - \xi_1} \tag{10.4}$$

The integrand contains the factor $\frac{1}{2} - n_F(\xi_1)$, and the $\frac{1}{2}$ term may be ignored. It does not cause the singularity, and its inclusion leads to little effect. The term in $n_F(\xi_1)$ leads to a logarithm which is singular at the Fermi energy:

$$\int_{-\omega_D}^{\omega_D} d\xi_1 \frac{n_F(\xi_1)}{\xi - \xi_1} = \int_{-\omega_D}^{0} \frac{d\xi_1}{\xi - \xi_1} = -\ln\left(\frac{\xi}{\omega_D}\right) \tag{10.5}$$

The results of the first and second Born approximation may be summarized as

$$V_{\text{eff}} = -V_0\left[1 - N_F V_0 \ln\left(\frac{\xi}{\omega_D}\right)\right] \tag{10.6}$$

The term $-N_F V_0 \ln(\xi/\omega_D)$ is regarded as the vertex correction which results from the additional scattering between electrons. The scattering becomes very large for electrons near the Fermi energy $|\xi/\omega_D| \ll 1$. Further insight is gained by considering the sum of diagrams in

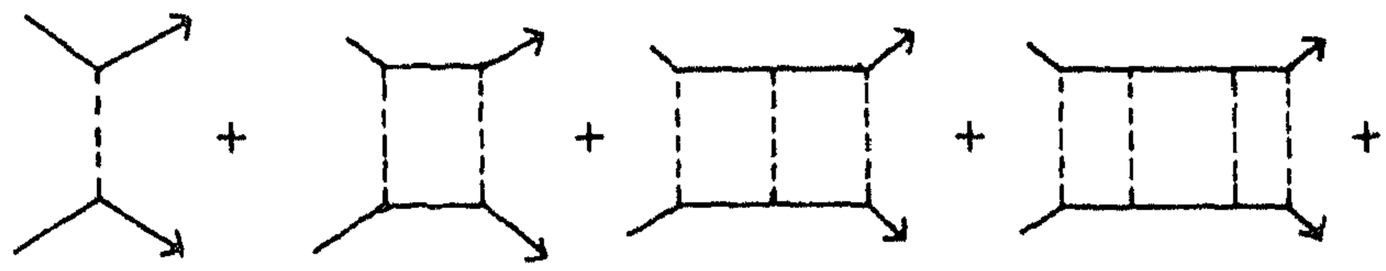


FIGURE 10.3

Fig. 10.3. Each additional interaction (dashed line) causes two more Green's functions which are going parallel in the intermediate state. Each new set of intermediate states has the same type of integrand, so that a term with $(n+1)$ ladder diagrams gives a net contribution of

$$-V_0\left[-N_F V_0 \ln\left(\frac{\xi}{\omega_D}\right)\right]^n \tag{10.7}$$

Each term with n ladder diagrams has a factor of $[-N_F V_0 \ln(\xi/\omega_D)]^n$, and the summation of these terms produces the series

$$V_{\text{eff}} = -V_0 \sum_{N=0}^{\infty}\left[-N_F V_0 \ln\left(\frac{\xi}{\omega_D}\right)\right]^n \tag{10.8}$$

$$= -\frac{V_0}{1 + N_F V_0 \ln(\xi/\omega_D)} \tag{10.9}$$

This series can be summed to produce a net potential with an energy denominator (10.9). This energy denominator equals zero at

$$\xi_0 = \omega_D \exp[-1/N_F V_0] \tag{10.10}$$

The energy denominator has a pole at this energy, since the denominator can be written as $N_F V_0 \ln(\xi/\xi_0)$. In the vicinity of ξ_0 the denominator can be approximated using $\xi = \xi_0 + (\xi - \xi_0)$ and the scattering potential has a pole:

$$V_{\text{eff}} = -\frac{1}{N_F}\frac{1}{\ln(\xi/\xi_0)} = -\frac{1}{N_F}\frac{1}{\ln[1 + (\xi - \xi_0)/\xi_0]}$$

$$\approx \frac{\xi_0}{N_F(\xi - \xi_0)} \tag{10.11}$$

This pole is sufficient to cause the instability. The electrons near the Fermi energy will interact with their pair on the opposite side of the Fermi sea. The mutual scattering produces a pole in the scattering amplitude, which will make the pair of electrons try to bind together. Of course, all electron pairs are doing this simultaneously, so that the entire metal undergoes a phase transition. The existence of this pole depended on the sharpness of the electron distribution $n_F(\xi_1) = \Theta(-\xi_1)$. If all the electrons near the Fermi energy become paired, one must reconsider whether this sharp distribution still exists. A theory of superconductivity must self-consistently determine the properties of bound electron pairs. This derivation is the basis of the BCS theory.

Another way to describe the instability is as a function of temperature. The calculation is repeated as a function of temperature. It enters into the electron distribution $n_F(\xi_1)$ by

changing the step $\Theta(-\xi_1)$ into a smooth function with an energy width of several $k_B T$. The influence of $n_F(\xi_1)$ can be approximated by expressing the integral (10.5) as

$$\int_{-\omega_D}^{\omega_D} d\xi_1 \frac{n_F(\xi_1)}{\xi - \xi_1} \approx -\frac{1}{2}\ln\left[\frac{\xi^2 + (k_B T)^2}{\omega_D^2}\right] \tag{10.12}$$

Follow the same steps through the summation of all the ladder diagrams, and conclude that the effective potential now has an energy denominator of the form

$$V_{\text{eff}} = -\frac{V_0}{1 + N_F V_0 \ln[\sqrt{\xi^2 + (k_B T)^2}/\omega_D]} \tag{10.13}$$

At zero energy ($\xi = 0$), V_{eff} becomes singular when the temperature is lowered to the critical temperature T_c:

$$k_B T_c = \omega_D \exp\left[-\frac{1}{N_F V_0}\right] \tag{10.14}$$

This argument correctly predicts the right form for the transition temperature of the BCS theory, which is

$$k_B T_c = 1.14\omega_D \exp\left[-\frac{1}{N_F V_0}\right] \tag{10.15}$$

The result is indeed proportional to the Debye energy, in agreement with the isotope effect. Of course, to explain the isotope effect, it must be shown that the exponent $1/(N_F V_0)$ does not change with the ion mass. It actually does depend on the ion mass M, which is contained in the electron–phonon matrix element, in the interaction potential V_0. Some experimental values of T_c for elemental superconductors are given in Table 10.1.

The theory of the Cooper instability should be compared, for example, with the ordinary binding of two isolated particles. If the two particles are isolated, they do not have to obey the statistics of a collection of identical particles. Then the scattering theory does not contain any of the occupation factors; all states may be used as intermediate states since there are no other particles. In this case the multiple scattering theory was described earlier, in Sec. 9.2, in the theory of Wannier excitons. The multiple scattering theory may be described by a vertex function:

$$\Gamma(\mathbf{k}, \mathbf{k}') = V(\mathbf{k} - \mathbf{k}') + \int \frac{d^3 k_1}{(2\pi)^3} \frac{V(\mathbf{k} - \mathbf{k}_1)\Gamma(\mathbf{k}_1, \mathbf{k}')}{2\xi_{\mathbf{k}} - 2\xi_{\mathbf{k}_1}} \tag{10.16}$$

TABLE 10.1 Critical temperatures T_c of elemental superconductors (K).

Mg	Al	
—	1.2	
Zn	Ga	
0.91	1.1	
Cd	In	Sn
0.56	3.37	3.73
Hg	Tl	Pb
4.16	2.38	7.22

The solution to this vertex function is equivalent to solving the two-particle Schrödinger equation in relative coordinates:

$$\left[-\frac{1}{2m}(\nabla_1^2+\nabla_2^2)+V(\mathbf{r}_1-\mathbf{r}_2)-E\right]\psi(\mathbf{r}_1,\mathbf{r}_2)=0 \tag{10.17}$$

The problem is factored into the relative and center of mass motions:

$$\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2, \qquad \psi(\mathbf{r}_1,\mathbf{r}_2)=e^{i\mathbf{p}\cdot\mathbf{R}}\phi(\mathbf{r}) \tag{10.18}$$

$$\mathbf{R}=\tfrac{1}{2}(\mathbf{r}_1+\mathbf{r}_2), \qquad E=\frac{p^2}{2m}+\varepsilon \tag{10.19}$$

$$\left[-\frac{\nabla^2}{m}+V(\mathbf{r})-\varepsilon\right]\phi(\mathbf{r})=0 \tag{10.20}$$

The center of mass motion is plane wave, and the relative motion becomes a one-body problem. Without the occupation factors, the relative scattering of two particles by an instantaneous potential is a trivial problem. When bound states occur, they are at negative binding energy in the relative coordinates, i.e., at $\varepsilon < 0$. This behavior is in great contrast to the Cooper instability, where the pole occurs at a small negative energy relative to the Fermi energy E_F, so the pole is at a positive energy $E_F - \xi_0$. Two electrons cannot really bind at that energy, since their net energy is positive. The instability occurs because it appears to them as if they should bind, although if they tried, they would find they could not. The role of the occupation factors $[1-2n_F(\xi_1)]$ in the argument of the scattering integral is what moved the apparent pole out to the Fermi energy.

One can now see the reason the electrons must be paired with opposite momentum. The instability is caused by the sharpness of the Fermi surface, which is fixed in momentum space. Two electrons with arbitrary wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ interact by exchanging phonons and this interaction can be attractive. Nevertheless, this does not lead to an instability when the center of mass motion is nonzero, since the zero energy of relative motion does not coincide with the location of the discontinuity in the momentum distribution. The center of mass transformation

$$\mathbf{P}=\mathbf{k}_1+\mathbf{k}_2, \qquad \mathbf{k}_1=\tfrac{1}{2}\mathbf{P}+\mathbf{k} \tag{10.21}$$

$$\mathbf{k}=\tfrac{1}{2}(\mathbf{k}_1-\mathbf{k}_2), \qquad \mathbf{k}_2=\tfrac{1}{2}\mathbf{P}-\mathbf{k} \tag{10.22}$$

$$\xi_{\mathbf{k}_1}+\xi_{\mathbf{k}_2}=\frac{k^2}{m}+\frac{P^2}{4m}-2E_F \tag{10.23}$$

changes the relative scattering integral of two particles into

$$\int\frac{d^3q}{(2\pi)^3}V(q)\frac{1-n_F(\xi_{\frac{1}{2}\mathbf{P}-\mathbf{k}-\mathbf{q}})-n_F(\xi_{\frac{1}{2}\mathbf{P}+\mathbf{k}+\mathbf{q}})}{k^2/m-(\mathbf{k}+\mathbf{q})^2/m} \tag{10.24}$$

The energy denominator factors into just the difference of the relative energies, as it should, but the Fermi surface discontinuities are not at $(\mathbf{k}+\mathbf{q})^2/m$. Instead, they are at $(\pm\frac{1}{2}\mathbf{P}+\mathbf{k}+\mathbf{q})^2/2m$, which gets smeared with the averaging over angles. This smearing makes the pole disappear rapidly for nonzero center of mass momentum $\mathbf{P}$.

The same results can also be derived from a Green's function analysis. If two particles interact with an instantaneous interaction, as in Fig. 10.2, and propagate parallel in time or τ space, then the vertex has the combination of Green's functions

$$\mathscr{G}^{(0)}(\mathbf{k}_1,\tau-\tau')\mathscr{G}^{(0)}(\mathbf{k}_2,\tau-\tau') \tag{10.25}$$

The frequency spectrum is obtained by integrating this pair over $(\tau - \tau')$

$$X(i\omega) = \int_0^\beta d\tau e^{i\omega(\tau-\tau')}\mathscr{G}^{(0)}(\mathbf{k}_1, \tau - \tau')\mathscr{G}^{(0)}(\mathbf{k}_2, \tau - \tau') \tag{10.26}$$

This integral can be evaluated directly in τ space by using the definitions of $\mathscr{G}^{(0)}$ in Sec. 3.2. Alternately, it can be converted to a frequency summation, which has the form

$$X(i\omega) = \frac{1}{\beta}\sum_{ip}\mathscr{G}^{(0)}(\mathbf{k}_1, ip)\mathscr{G}^{(0)}(\mathbf{k}_2, i\omega - ip) \tag{10.27}$$

$$= \frac{1 - n_F(\xi_{\mathbf{k}_1}) - n_F(\xi_{\mathbf{k}_2})}{i\omega - \xi_{\mathbf{k}_1} - \xi_{\mathbf{k}_2}} \tag{10.28}$$

The expression contains one of the standard Matsubara summations which was given in Sec. 3.5. It has exactly the combination of occupation number factors stated in (10.24). It also has the same energy denominators. The binding of particles is described by parallel motion in time.

It is interesting to generalize the Cooper instability to other circumstances in which similar effects could occur. For example, consider the interaction between two different fermions, e.g., electrons and holes in a semimetal or semiconductor. They have the Fermi distributions $n_e(\xi_k)$ and $n_h(\xi_k)$, respectively. The scattering theory contains the vertex term for the scattering $(\mathbf{k}_e, \mathbf{k}_h) \to (\mathbf{k}_e + \mathbf{q}, \mathbf{k}_h - \mathbf{q})$.

$$\int \frac{d^3q}{(2\pi)^3} V(q) \frac{1 - n_e(\mathbf{k}_e + \mathbf{q}) - n_F(\mathbf{k}_h - \mathbf{q})}{\varepsilon_{e,\mathbf{k}_e} + \varepsilon_{h,\mathbf{k}_h} - \varepsilon_{e,\mathbf{k}_e+\mathbf{q}} - \varepsilon_{k,\mathbf{k}_h-\mathbf{q}}} \tag{10.29}$$

This vertex would also cause V_{eff} to be resonant if one had $\mathbf{k}_h = -\mathbf{k}_e$, but this cannot happen in a solid in equilibrium, since otherwise the electrons and holes would recombine. However, one can still get instabilities out of this process, as was originally discussed by Keldysh and Kopaev (1965). The instability leads to a state called the *excitonic insulator*, which has been reviewed by Halperin and Rice (1968).

A logarithmic singularity is obtained in (10.29) even if one of the occupation factors is zero, say $n_h = 0$, and only one particle obeys many-particle statistics. One example is the edge singularities in X-ray spectra first predicted by Mahan (1967) and discussed in Sec. 9.3.2. In electron–hole scattering, the electron occupation factors remain in the scattering function even if $n_h = 0$

$$(1 - n_e)(1 - n_h) - n_e n_h = 1 - n_e - n_h \approx 1 - n_e \tag{10.30}$$

This behavior is in sharp contrast to the multiple scattering theory for an electron scattering from a fixed potential such as impurity scattering. Then there is only one fermion, say an electron, which has a factor $(1 - n_e)$ for scattering $\mathbf{k} \to \mathbf{k}'$ and the factor $(-n_e)$ for the rate of back scattering $\mathbf{k}' \to \mathbf{k}$. The back scattering enters with a sign change because of the anti-symmetry of the single fermion state. The two rates are added, $1 - n_e + n_e = 1$, and the occupation number terms n_e cancel out. The electron scattering from an impurity is treated as a one-body problems (except for self-consistent screening) even in a many-particle system. This cancellation explains why such factors as $n_F(\xi)$ do not appear in the vertex equations for impurity scattering, which were solved in Sec. 8.1.

Another possibility is to have one or both of the particles be bosons. Then the scattering probability is multiplied by $[1 + n_B(\xi_1)]$ for scattering into state $\mathbf{k}_1$ and by $n_B(\xi_1)$ for scattering out of it. For a single boson scattering from a fixed potential, the scattering integral has

the rate $(1 + n_B)$ for $\mathbf{k} \to \mathbf{k}_1$ and the rate n_B back, so the net is $1 + n_B - n_B = 1$, which does not contain occupation factors. The potential scattering is again a one-body problem even in a many-particle system. However, two bosons which are mutually scattering have the factors $(1 + n_B)(1 + n'_B) - n_B n'_B = 1 + n_B + n'_B$, so that the scattering is enhanced because of the occupation of other particles.

10.1.1. BCS Theory

The basic feature of the BCS theory is that pairing occurs between electrons in states with opposite momentum and opposite spins, e.g., between states $(\mathbf{k}, \uparrow)$ and $(-\mathbf{k}, \downarrow)$. The two spins are combined into a spin singlet, with $S = 0$. The singlet was chosen in BCS theory on the basis that the other choices of spin combination would lead to a triplet state with $S = 1$. The latter choice $(S = 1)$ implies the superconducting state has magnetic properties, which in fact are absent for simple metals. The choice $S = 0$ seems most reasonable. Later work by Balian and Werthamer (1963), who solved the BCS equations for $S = 1$, showed that the triplet state had smaller binding energy and was therefore less favored. However, the recent theories of superfluidity in ^{3}He are based on the premise that the pairing occurs in the triplet state (see Chapter 11). Triplet pairing is possible and may exist in heavy fermion solids such as UPt_3 and UBe_{13} (Han *et al.*, 1986). The discussion will assume that the spin arrangement is singlet.

The pairing of electrons in the BCS theory must cause correlations in their relative motions. The pairing is described by introducing a new correlation function, similar to a Green's function, for particles of opposite spin. Following Abrikosov *et al.* (1963), these are

$$\begin{aligned} \mathscr{G}(\mathbf{p}, \tau - \tau') &= -\langle T_\tau C_{\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle \\ \mathscr{F}(\mathbf{p}, \tau - \tau') &= \langle T_\tau C_{-\mathbf{p}\downarrow}(\tau) C_{\mathbf{p}\uparrow}(\tau')\rangle \\ \mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') &= \langle T_\tau C^\dagger_{\mathbf{p}\uparrow}(\tau) C^\dagger_{-\mathbf{p}\downarrow}(\tau')\rangle \end{aligned} \tag{10.31}$$

The Green's function $\mathscr{G}$ has the same definition as usual, although it has a different algebraic form in the superconducting state. The $\mathscr{F}$ and $\mathscr{F}^\dagger$ functions are identically zero in the normal state. Even in the superconducting state, the bra $\langle|$ and ket $|\rangle$ notation must have special meaning. For instance, in the definition of the $\mathscr{F}$ function, the ket $|\rangle$ is operated on by two destruction operators and then closed by $\langle|$. The state $|\rangle$ must have two more electrons than $\langle|$. A basic feature of the BCS ground state wave function is that it is composed of a superposition of electronic states containing a different number of electrons. Such a formulation is possible in a grand canonical ensemble. The procedure will be to find a self-consistent equation for the correlation function $\mathscr{F}$ or its Hermitian conjugate $\mathscr{F}^\dagger$. At high temperatures the only solution is $\mathscr{F} = \mathscr{F}^\dagger = 0$, but a nonzero solution becomes possible at low temperature.

The order of the spin indices $(\uparrow, \downarrow)$ is important in specifying the Green's functions. Suppose by mistake for $\mathscr{F}$ one wrote

$$\langle T_\tau C_{-\mathbf{p}\uparrow}(\tau) C_{\mathbf{p}\downarrow}(\tau')\rangle \tag{10.32}$$

Since $\mathscr{F}$ does not depend on the sign of $\mathbf{p}$, this correlation function is the same as

$$\langle T_\tau C_{\mathbf{p}\uparrow}(\tau) C_{-\mathbf{p}\downarrow}(\tau')\rangle \tag{10.33}$$

The $C_{\mathbf{p}\sigma}$ operators anticommute, and doing this gives

$$-\langle T_\tau C_{-\mathbf{p}\downarrow}(\tau')C_{\mathbf{p}\uparrow}(\tau)\rangle \tag{10.34}$$

$-\mathscr{F}(\mathbf{p}, \tau' - \tau)$. These kinds of sign errors can be avoided by always following the definition closely.

The objective is to provide the simplest possible derivation of the BCS theory. First assume a model Hamiltonian of the form

$$H = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \frac{1}{2\nu} \sum_{\mathbf{q}\mathbf{p}\mathbf{p}'ss'} V(q) C^\dagger_{\mathbf{p}+\mathbf{q},s} C^\dagger_{\mathbf{p}'-\mathbf{q},s'} C_{\mathbf{p}'s'} C_{\mathbf{p}s} \tag{10.35}$$

The interaction potential $V(q)$ between electrons is taken to have the form in (10.2), which is an attractive constant $V(q) = -V_0$ over a range of energies within a Debye energy of the Fermi surface. With this Hamiltonian, a set of self-consistent equations will be derived for the Green's functions $\mathscr{G}$, $\mathscr{F}$, and $\mathscr{F}^\dagger$. The derivation is done using the equation of motion. Consider

$$\frac{d}{d\tau} C_{\mathbf{p}\sigma}(\tau) = [H, C_{\mathbf{p}\sigma}] = -\xi_{\mathbf{p}} C_{\mathbf{p}\sigma} = \frac{1}{\nu} \sum_{\mathbf{q}\mathbf{p}'s'} V(q) C^\dagger_{\mathbf{p}'-\mathbf{q},s'} C_{\mathbf{p}'s'} C_{\mathbf{p}-\mathbf{q},\sigma} \tag{10.36}$$

From the definition of the τ-ordered product, the first derivative of the equation for the Green's function is:

$$\frac{\partial}{\partial\tau} \mathscr{G}(\mathbf{p}, \tau - \tau') = -\frac{\partial}{\partial\tau}[\Theta(\tau - \tau')\langle C_{\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle - \Theta(\tau' - \tau)\langle C^\dagger_{\mathbf{p}\sigma}(\tau') C_{\mathbf{p}\sigma}(\tau)\rangle] \tag{10.37}$$

$$= -\delta(\tau - \tau') - \left\langle T_\tau \left[\frac{\partial}{\partial\tau} C_{\mathbf{p}\sigma}(\tau)\right] C^\dagger_{\mathbf{p}\sigma}(\tau')\right\rangle \tag{10.38}$$

Using the result (10.36) for $dC_{\mathbf{p}\sigma}/d\tau$ gives

$$\left(-\frac{\partial}{\partial\tau} - \xi_p\right)\mathscr{G}(\mathbf{p}, \tau - \tau') + \frac{1}{\nu} \sum_{\mathbf{q}\mathbf{p}'s} V(q) \tag{10.39}$$

$$\times \langle T_\tau C^\dagger_{\mathbf{p}'-\mathbf{q},s'}(\tau) C_{\mathbf{p}'s'}(\tau) C_{\mathbf{p}-\mathbf{q},\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle = \delta(\tau - \tau') \tag{10.40}$$

The bracket of four operators in the interaction term must be reduced to products of pair operators. There are many ways of doing the pairing, since in addition to the normal combinations such as $\langle CC^\dagger\rangle\langle CC^\dagger\rangle$, there are also arrangements such as $\langle CC\rangle\langle C^\dagger C^\dagger\rangle$. One simplification is to assume that long-wavelength phonons give a zero potential, so that $V(\mathbf{q} = 0) = 0$. Neglect the pairing which occurs when $\mathbf{q} = 0$. For a normal metal, there would only remain the pairing $\delta_{\mathbf{p}=\mathbf{p}'}\delta_{s'=\sigma} n_{\mathbf{p}-\mathbf{q}} \mathscr{G}(\mathbf{p}, \tau - \tau')$ which gives the exchange energy. This pairing occurs in the superconductor as well but is not the only term. The pairings which include the $\mathscr{F}$ functions must pay attention to the spin variables. The combination $\sigma = -s' = \uparrow$ gives

$$\langle T_\tau C^\dagger_{\mathbf{p}'-\mathbf{q},s'}(\tau) C_{\mathbf{p}'s'}(\tau) C_{\mathbf{p}-\mathbf{q},\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle \tag{10.41}$$

$$= -\delta_{\sigma=-s'}\delta_{\mathbf{p}'=-\mathbf{p}+\mathbf{q}} \mathscr{F}(\mathbf{p} - \mathbf{q}, 0)\mathscr{F}^\dagger(\mathbf{p}, \tau' - \tau) \tag{10.42}$$

where the sign change resulted from an odd number of operator rearrangements. Similarly, the choice $\sigma = -s' = \downarrow$ gives

$$\langle T_\tau C^\dagger_{\mathbf{p}'-\mathbf{q},s'}(\tau) C_{\mathbf{p}'s'}(\tau) C_{\mathbf{p}-\mathbf{q},\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle \tag{10.43}$$

$$= -\delta_{\sigma=-s'}\delta_{\mathbf{p}'=-\mathbf{p}+\mathbf{q}}\mathscr{F}(-\mathbf{p}+\mathbf{q},0)\mathscr{F}^\dagger(-\mathbf{p},\tau-\tau') \tag{10.44}$$

These two results are identical, since later it is shown that the $\mathscr{F}$ and $\mathscr{F}^\dagger$ functions do not depend on the sign of their arguments either momentum or τ. The last term in (10.40) gives the expression

$$\frac{1}{\nu}\sum_{\mathbf{q}\mathbf{p}'s} V(q)\langle T_\tau C^\dagger_{\mathbf{p}'-\mathbf{q},s'}(\tau) C_{\mathbf{p}'s'}(\tau) C_{\mathbf{p}-\mathbf{q},\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle$$

$$= \frac{1}{\nu}\sum_{\mathbf{q}} V(q)[\mathscr{G}(\mathbf{p},\tau-\tau')n_{\mathbf{p}-\mathbf{q}} - \mathscr{F}(\mathbf{p}-\mathbf{q},0)\mathscr{F}^\dagger(\mathbf{p},\tau-\tau')] \tag{10.45}$$

The first term is the exchange self-energy of the electron due to the phonon induced interaction between electrons. It provides the self-energy term $\Sigma_x = -\Sigma V(q)n_{\mathbf{p}-\mathbf{q}}$. A careful investigation shows that this self-energy does not change much between the normal and superconducting states. The self-energy of the electrons, from phonons, causes a change in the electron effective mass given by the parameter λ. This effect is not large in weak superconductors, so it may be ignored. In metals where λ is large, such as lead, the superconducting state can be expected to significantly alter the properties of electrons near the Fermi surface. Hence the need, in these cases, for strong coupling theory. This self-energy term is neglected in our weak coupling theory.

In the second term of (10.45) there arises the combination of factors which are defined as

$$\Delta(\mathbf{p}) = -\frac{1}{\nu}\sum_{\mathbf{q}} V(q)\mathscr{F}(\mathbf{p}-\mathbf{q},\tau=0) \tag{10.46}$$

The quantity $\Delta(\mathbf{p})$ is the gap function in the BCS theory and plays a central role in the properties of the superconductor state. The quantity $\Delta(\mathbf{p})$ is defined to be positive, since the right-hand side of the definition is positive, with an attractive potential $V(q) < 0$. Collecting these results in (10.40) gives

$$\left(-\frac{\partial}{\partial\tau} - \xi_p\right)\mathscr{G}(\mathbf{p},\tau-\tau') + \Delta(\mathbf{p})\mathscr{F}^\dagger(\mathbf{p},\tau-\tau') = \delta(\tau-\tau') \tag{10.47}$$

The equation has two unknowns in $\mathscr{G}$ and $\mathscr{F}^\dagger$, so another equation is needed to link these two quantities. It comes from the equation of motion for the $\mathscr{F}^\dagger$ function:

$$\frac{\partial}{\partial\tau}\mathscr{F}^\dagger(\mathbf{p},\tau-\tau') = -\frac{\partial}{\partial\tau}[\Theta(\tau-\tau')\langle C^\dagger_{\mathbf{p}\uparrow}(\tau)C^\dagger_{-\mathbf{p}\downarrow}(\tau')\rangle \tag{10.48}$$

$$- \Theta(\tau'-\tau)\langle C^\dagger_{-\mathbf{p}\downarrow}(\tau')C^\dagger_{\mathbf{p}\uparrow}(\tau)\rangle]$$

$$= -\delta(\tau-\tau')\langle\{C^\dagger_{\mathbf{p}\uparrow}, C^\dagger_{-\mathbf{p}\downarrow}\}\rangle + \left\langle T_\tau\left[\frac{\partial}{\partial\tau}C^\dagger_{\mathbf{p}\uparrow}(\tau)\right]C^\dagger_{-\mathbf{p}\downarrow}(\tau')\right\rangle$$

$$= \left\langle T_\tau\left[\frac{\partial}{\partial\tau}C^\dagger_{\mathbf{p}\uparrow}(\tau)\right]C^\dagger_{-\mathbf{p}\downarrow}(\tau')\right\rangle \tag{10.49}$$

It lacks a term $\delta(\tau - \tau')$ because the $C^\dagger$ operators anticommute. The time development of the $\mathscr{F}^\dagger$ operator is determined by

$$\frac{d}{d\tau} C^\dagger_{\mathbf{p}\sigma}(\tau) = [H, C^\dagger_{\mathbf{p}\sigma}] = \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} + \frac{1}{\mathrm{V}} \sum_{\mathbf{q}\mathbf{p}'s'} V(q) C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C_{\mathbf{p}'+\mathbf{q},s'} C_{\mathbf{p}',s'}$$

This equation is not the Hermitian conjugate of $dC_{\mathbf{p}\sigma}(\tau)/d\tau$ in (10.36), because $C^\dagger_{\mathbf{p}\sigma}(\tau)$ is not the Hermitian conjugate of $C_{\mathbf{p}\sigma}(\tau)$. The result for $dC^\dagger/d\tau$ brings us to the equation for $\partial\mathscr{F}^\dagger/\partial\tau$:

$$\left(-\frac{\partial}{\partial\tau} + \xi_p\right)\mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') + \frac{1}{\mathrm{V}} \sum_{\mathbf{q}\mathbf{p}'s} V(q) \tag{10.50}$$

$$\times \langle T_\tau C^\dagger_{\mathbf{p}-\mathbf{q},\uparrow}(\tau) C^\dagger_{\mathbf{p}'+\mathbf{q},s'}(\tau) C_{\mathbf{p}'s'}(\tau) C^\dagger_{-\mathbf{p}\downarrow}(\tau')\rangle = 0 \tag{10.51}$$

Again there is an expression for four operators. The operator $C^\dagger_{-\mathbf{p}\downarrow}(\tau')$ is unique, since it is the only one which is not operating at time τ. Pairing it with each of the other three operators gives three terms:

$$\begin{aligned} C_{\mathbf{p}'s'}(\tau)&: \quad \delta_{s'=\downarrow} d_{\mathbf{p}'=-\mathbf{p}} \Delta^\dagger(\mathbf{p}) \mathscr{G}(-\mathbf{p}, \tau - \tau') \\ C^\dagger_{\mathbf{p}-\mathbf{q}\uparrow}(\tau)&: \quad \delta_{\mathbf{q}=0} n_{\mathbf{p}'} \mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') \\ C^\dagger_{\mathbf{p}'+\mathbf{q},s'}(\tau)&: \quad -\delta_{s'=\uparrow} \delta_{\mathbf{p}'=\mathbf{p}-\mathbf{q}} n_{\mathbf{p}-\mathbf{q}} \mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') \end{aligned} \tag{10.52}$$

The first term has $\Delta^\dagger(\mathbf{p})$, which is the important term; the second requires $\mathbf{q} = 0$, which was assumed is zero; and the last gives the exchange potential $\Sigma_x = -\sum V(q) n_{\mathbf{p}+\mathbf{q}}$, which is ignored again. The exchange potential just renormalizes the chemical potential. This term is omitted. These steps brings us to the final equation:

$$\left(-\frac{\partial}{\partial\tau} + \xi_p\right)\mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') + \Delta^\dagger(\mathbf{p}) \mathscr{G}(\mathbf{p}, \tau - \tau') = 0 \tag{10.53}$$

The gap function is assumed real, $\Delta^\dagger(\mathbf{p}) = \Delta(\mathbf{p})$, which is verified later.

The Fourier transforms of the correlation functions are defined in the usual way:

$$\begin{aligned} \mathscr{G}(\mathbf{p}, \tau) &= \frac{1}{\beta} \sum_{p_n} e^{-ip_n\tau} \mathscr{G}(\mathbf{p}, ip_n) \\ \mathscr{F}(\mathbf{p}, \tau) &= \frac{1}{\beta} \sum_{p_n} e^{-ip_n\tau} \mathscr{F}(\mathbf{p}, ip_n) \\ \mathscr{F}^\dagger(\mathbf{p}, \tau) &= \frac{1}{\beta} \sum_{p_n} e^{-ip_n\tau} \mathscr{F}^\dagger(\mathbf{p}, ip_n) \end{aligned} \tag{10.54}$$

After transforming, the two equations (10.47) and (10.53) are

$$(ip_n - \xi_p)\mathscr{G}(\mathbf{p}, ip_n) + \Delta(\mathbf{p})\mathscr{F}^\dagger(\mathbf{p}, ip_n) = 1 \tag{10.55}$$

$$(ip_n + \xi_p)\mathscr{F}^\dagger(\mathbf{p}, ip_n) + \Delta(\mathbf{p})\mathscr{G}(\mathbf{p}, ip_n) = 0 \tag{10.56}$$

These algebraic equations are easily solved:

$$\mathscr{G}(\mathbf{p}, ip_n) = -\frac{ip_n + \xi_p}{p_n^2 + \xi_p^2 + \Delta^2(\mathbf{p})} \tag{10.57}$$

$$\mathscr{F}(\mathbf{p}, ip_n) = \mathscr{F}^\dagger(\mathbf{p}, ip_n) = \frac{\Delta(\mathbf{p})}{p_n^2 + \xi_p^2 + \Delta^2(\mathbf{p})} \tag{10.58}$$

The equivalence of $\mathscr{F}$ and $\mathscr{F}^\dagger$ can be shown by deriving similar equations for $\mathscr{F}$.

The results (10.58) can be used to test some of the assumptions in the derivation. For example, consider the Fourier transform (10.45) of $\mathscr{F}(\mathbf{p}, -\tau)$. Changing the sign of the dummy variable of summation gives

$$\mathscr{F}(\mathbf{p}, -\tau) = \frac{1}{\beta}\sum_n e^{-i\tau p_n}\mathscr{F}(\mathbf{p}, -ip_n) = \mathscr{F}(\mathbf{p}, \tau) \tag{10.59}$$

Similarly,

$$\mathscr{F}(\mathbf{p}, \tau = 0) = \frac{1}{\beta}\sum_n \mathscr{F}(\mathbf{p}, ip_n) = \frac{1}{\beta}\sum_n \mathscr{F}^\dagger(\mathbf{p}, ip_n) = \mathscr{F}^\dagger(\mathbf{p}, \tau = 0)$$

This relation causes the energy gap equation to be real, and for $\Delta^\dagger(\mathbf{p}) = \Delta(\mathbf{p})$

$$\begin{aligned}\Delta(\mathbf{p}) &= -\frac{1}{\nu}\sum_{\mathbf{q}} V(q)\mathscr{F}(\mathbf{p}-\mathbf{q}, 0) = -\frac{1}{\nu}\sum_{\mathbf{q}} V(q)\mathscr{F}^\dagger(\mathbf{p}-\mathbf{q}, 0)\\ &= \Delta^\dagger(\mathbf{p})\end{aligned} \tag{10.60}$$

Note that if the energy gap $\Delta(\mathbf{p})$ is set equal to zero, there is found the usual form for the normal state Green's function and zero for the others:

$$\Delta = 0: \quad \begin{cases} \mathscr{G}(\mathbf{p}, ip) = \dfrac{1}{ip - \xi_p} \\ \mathscr{F} = \mathscr{F}^\dagger = 0 \end{cases} \tag{10.61}$$

It must still be shown that there is a self-consistent solution to the equations for $\Delta \neq 0$.

The Green's functions $\mathscr{G}$, $\mathscr{F}$ and $\mathscr{F}^\dagger$ have poles at the points $\pm E_p = \pm\sqrt{\xi_p^2 + \Delta(\mathbf{p})^2}$. E_p is called the *excitation energy* of the superconductor. It will occur repeatedly in the various formulas. In BCS theory the gap function $\Delta(\mathbf{p})$ for s-wave superconductors is not dependent on momentum $\mathbf{p}$ and is treated as a constant for each temperature. Of course, it very much depends on temperature and vanishes at the transition temperature of the superconductor. However, Δ is treated as a constant at a fixed temperature. [Later, in strong coupling theory, $\Delta(i\omega_n)$ is found to be a function of energy.] The excitation energy E_p depends on $\xi_p^2 > 0$. The minimum excitation energy is Δ, and $E_p \geq \Delta$. An important feature of the BCS theory is that the particles are paired, and it is not possible to excite just one quasiparticle with excitation energy E_p. Instead, one must break a pair of particles and excite them both to the band of excitations. This pair breaking is shown in Fig. 10.4, where the horizontal line represents the bound state pairs at the chemical potential of the superconductor. To break a pair, one must excite both particles to the excitation line, so that it takes energy $E_p + E_{p'} \geq 2\Delta$. The minimum excitation energy of the superconductor is the energy to break a pair, which is $E_g \equiv 2\Delta$, where E_g is the energy gap of the superconductor. This fact must be kept in mind when comparing thermodynamic data with the BCS theory.

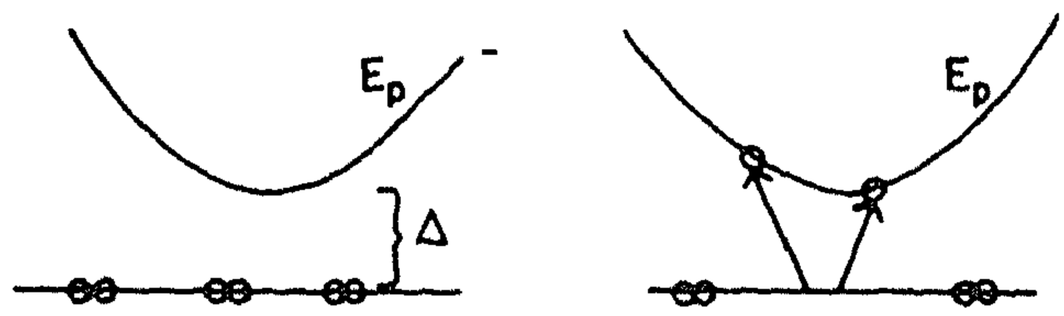


FIGURE 10.4 BCS model of superconductor. (a) The ground state has electrons paired at the chemical potential. (b) Excited states occur by breaking a pair and forming two quasiparticles, each with excitation energy E_p.

The existence of condensed pairs also clarifies the meaning of the bra $\langle|$ and ket $|\rangle$ symbols in the definition of

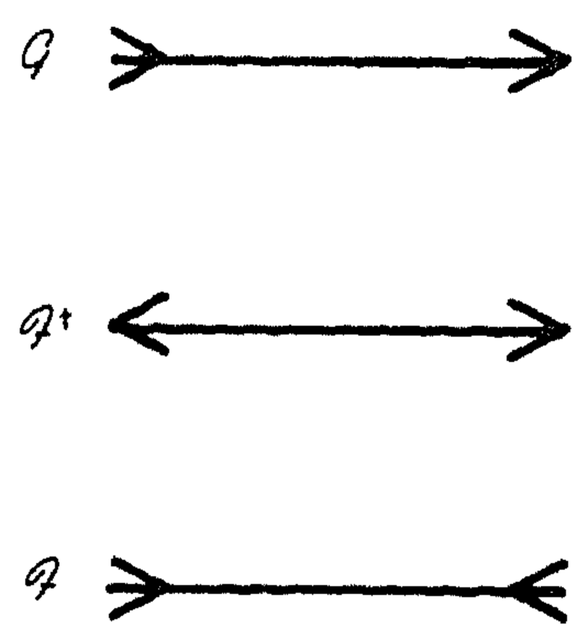
$$\mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') = \langle| T_\tau C^\dagger_{\mathbf{p}\uparrow}(\tau) C^\dagger_{-\mathbf{p}\downarrow}(\tau') |\rangle \tag{10.62}$$

At zero temperature, the ket $|\rangle$ is the ground state of the superconductor, which has all bound pairs at the chemical potential μ. The operators $C^\dagger_{\mathbf{p}\uparrow}$ and $C^\dagger_{-\mathbf{p}\downarrow}$ create two excitations with energy $E_\mathbf{p} + E_{-\mathbf{p}} = 2E_\mathbf{p}$. The bra $\langle|$ state has the same number of particles but two less in the ground state.

The Feynman diagrams for the three Green's functions are shown in Fig. 10.5. A diagrammatic representation is useful when evaluating correlation functions. The way of doing the diagrams was suggested by Abrikosov *et al.* (1963). The Green's function $\mathscr{G}$ represents creating a particle at one point in time and destroying it at a later point in time. It is drawn as an arrow with points at both ends and both points are in the same direction. The inward point at one end symbolizes particle destruction, and the outward point on the other is particle creation. $\mathscr{F}^\dagger$ is represented by an arrow with both points outward, since a particle is created at both times. Similarly, the Green's function $\mathscr{F}$ is an arrow with both points inward, which represents particle destruction at both points in time.

It is conventional to introduce the following coherence factors:

$$\begin{aligned} u_p^2 &= \frac{1}{2}\left(1 + \frac{\xi_p}{E_p}\right) \\ v_p^2 &= \frac{1}{2}\left(1 - \frac{\xi_p}{E_p}\right) \\ 2u_p v_p &= \left(1 - \frac{\xi_p^2}{E_p^2}\right)^{1/2} = \frac{\Delta}{E_p} \\ u_p^2 + v_p^2 &= 1 \end{aligned} \tag{10.63}$$



which are misnamed since they have little to do with coherence. Their usefulness stems from the fact that they are the residues of the poles of the Green's functions. Write the Green's functions (10.58) as

$$\mathscr{G}(\mathbf{p}, ip) = \frac{u_p^2}{ip - E_\mathbf{p}} + \frac{v_p^2}{ip + E_\mathbf{p}} \tag{10.64}$$

$$\mathscr{F}(\mathbf{p}, ip) = \mathscr{F}^\dagger(\mathbf{p}, ip) = -u_\mathbf{p} v_\mathbf{p}\left(\frac{1}{ip - E_\mathbf{p}} - \frac{1}{ip + E_\mathbf{p}}\right) \tag{10.65}$$

This form is useful for doing the contour integrals associated with the summations over Matsubara frequencies. The excitations of the superconductor are fermions, and the frequencies ip_n in $\mathscr{G}$ and $\mathscr{F}$ are for fermions $ip_n = (2n+1)\pi i/\beta$. The spectral function for the Green's function is

$$A(p, \varepsilon) = -2\,\mathrm{Im}[G_{\mathrm{ret}}(\mathbf{p}, \varepsilon)] = 2\pi[u_\mathbf{p}^2 \delta(\varepsilon - E_p) + v_\mathbf{p}^2 \delta(\varepsilon + E_p)] \tag{10.66}$$

These equations are not really solved until Δ is evaluated. It is found from the definition of the gap function in (10.46), where

$$\mathscr{F}(\mathbf{p}, \tau = 0) = \frac{1}{\beta}\sum_{p_n} \frac{\Delta}{p_n^2 + E_p^2} = \frac{\Delta}{2E_p}\tanh\left(\frac{\beta E_p}{2}\right) \tag{10.67}$$

The summation over frequencies is evaluated in the usual fashion, by the contour integral

$$\oint \frac{dz}{2\pi i} n_F(z) \frac{\Delta}{z^2 - E_p^2} = 0 \tag{10.68}$$

The integral is zero since the contour is taken to infinity. The poles of $n_F(z)$ give the summation over ip_n while the poles at $z = \pm E_p$ give the hyperbolic tangent result. The equation for the gap function is

$$\Delta(\mathbf{p}) = -\int \frac{d^3q}{(2\pi)^3} V(q) \frac{\Delta(\mathbf{p} - \mathbf{q})}{2E_{\mathbf{p}-\mathbf{q}}} \tanh\left(\frac{\beta E_{\mathbf{p}-\mathbf{q}}}{2}\right) \tag{10.69}$$

Since Δ is constant, it can be factored from both sides, which leaves the integral equation for Δ

$$1 = \frac{N_F V_0}{2} \int_{-\omega_D}^{\omega_D} \frac{d\xi}{E} \tanh\left(\frac{\beta E}{2}\right) \tag{10.70}$$

$$E = \sqrt{\xi^2 + \Delta^2} \tag{10.71}$$

The solution is found in two limiting cases. The first is at zero temperature, where the hyperbolic function $\tanh(\beta E/2) = 1$

$$1 = \frac{N_F V_0}{2} \int_{-\omega_D}^{\omega_D} \frac{d\xi}{\sqrt{\xi^2 + \Delta^2}} \tag{10.72}$$

$$= N_F V_0 \ln[\xi + \sqrt{\xi^2 + \Delta^2}]_0^{\omega_D} \approx N_F V_0 \ln\left(\frac{2\omega_D}{\Delta}\right) \tag{10.73}$$

which may be solved to produce the equation for the energy gap

$$E_g = 2\Delta = 4\omega_D \exp\left(-\frac{1}{N_F V_0}\right) \tag{10.74}$$

In BCS theory the gap equation (10.70) was solved as a function of temperature. The energy gap gets smaller as the temperature is increased, as shown in Fig. 10.6. In BCS theory the critical temperature was the result stated earlier in (10.15):

$$k_B T_c = 1.14\omega_D \exp\left(-\frac{1}{N_F V_0}\right) \tag{10.75}$$

The ratio of these two results predicts

$$\frac{E_g}{k_B T_c} = \frac{4.0}{1.14} = 3.52 \tag{10.76}$$

Both the energy gap E_g and the transition temperature T_c can be measured, so this can be tested. It is found to work well in weak superconductors such as aluminum, which is generally well described by the BCS theory. For strong coupling superconductors, the ratio increases in value.

The BCS theory is often called a mean field theory, because the gap equation (10.69) has that form. It has a similar form to the mean field theory for magnetism in metals (see Kittel, 1966; he uses the term *molecular field approximation*). The order parameter, in this case the gap Δ, is set equal to a thermodynamic average over the excitations, which are also functions of Δ. The self-consistent solution determines Δ. There is a critical temperature T_c above which no solution is possible. It is seldom that the mean field theory is an accurate description

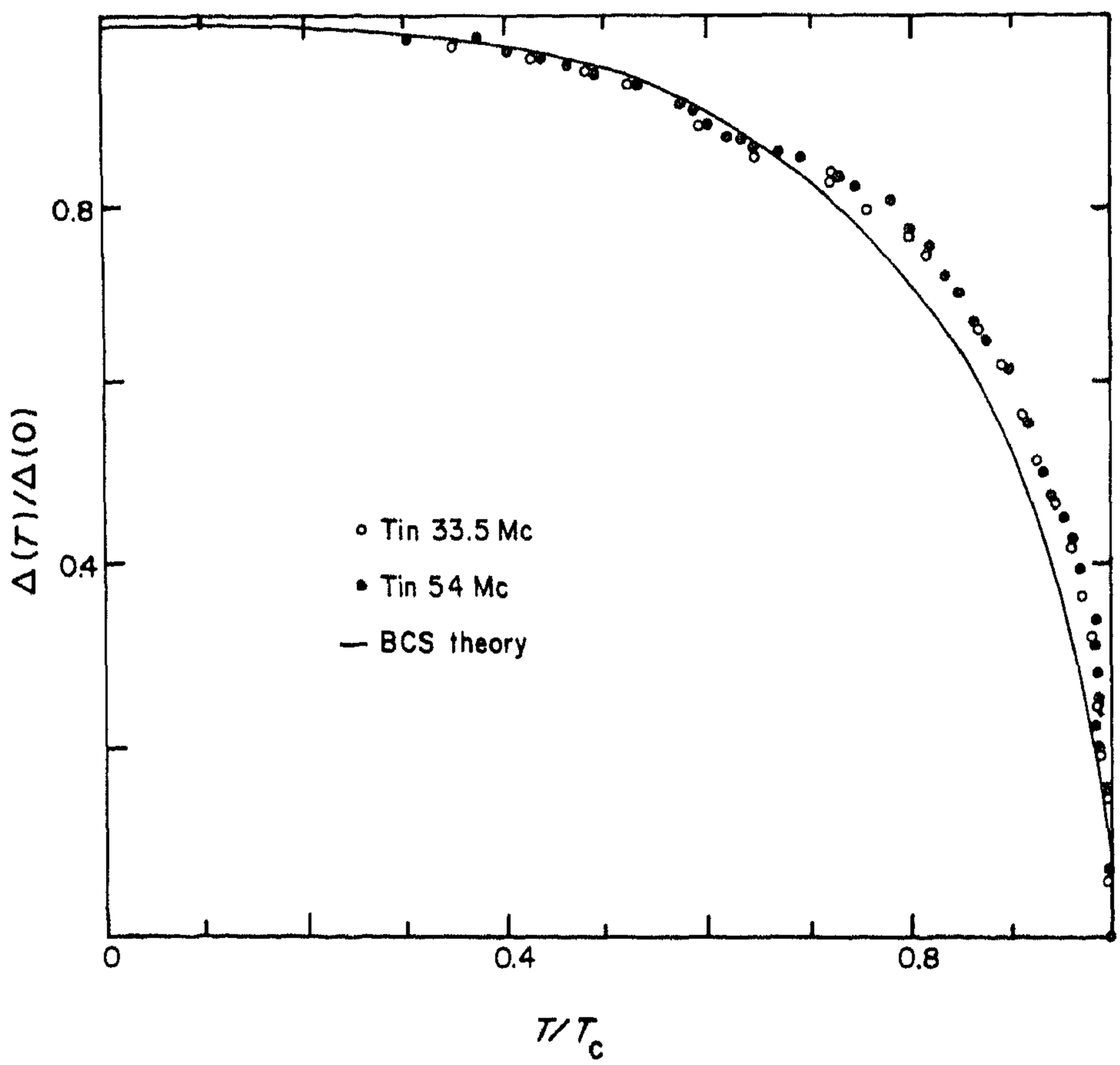


FIGURE 10.6 The ratio $\Delta(T)/\Delta(0)$ vs. T/T_c in tin. The solid curve is the BCS theory. The points are the ultrasonic attenuation data of Morse and Bohm (1957) (used with permission).

of physical reality, but it certainly works well for BCS. In statistical mechanics, mean field theory works well when the forces between particles are long range. In the case of superconductivity, the forces are short range, but the bound state orbits of the particles extend over a long distance.

Besides the existence of an energy gap, another dramatic prediction of the BCS theory was that the density of states $\rho(E)$ for excitations has a square root singularity. It is defined as

$$\rho(E) = \left(\frac{dE}{d\xi}\right)^{-1} = \frac{E}{\sqrt{E^2 - \Delta^2}} \tag{10.77}$$

This spectral shape is shown in Fig. 10.7, with a gap until Δ and then a square root singularity. The density of states arises from the variable change

$$\int_0^\infty d\xi = \int dE\left(\frac{d\xi}{dE}\right) = \int_\Delta^\infty dE\rho(E) \tag{10.78}$$

This density of states should not be regarded as just a function which results from a change of variables but as a real property of superconductors which can be measured.

The above theory is valid for a superconductor with an isotropic energy gap. Electron states in crystals belong to representations of the group which leaves the crystal invariant. Angular momentum eigenstates, such as s-, p-, etc., do not belong to representations of crystal groups. Instead, the crystal representations are mixtures of angular momentum states. The energy gap in crystals is found to vary somewhat with direction of the wave vector, because of this mixture of angular momentum states. However, the phrase "s-wave superconductivity" is used to describe the state which belongs to the most isotropic representation. In a cubic crystal, the gap would have the same value along all $\pm x$, $\pm y$, $\pm z$ axes. Similarly, the phrase "p-wave superconductivity" applies to crystals in which order parameters might have different signs in the $\pm z$ directions.

The cuprate superconductors may have "d-wave symmetry". The conduction electrons are confined to move in the (x, y) planes composed of copper and oxygen atoms. In some cases the x and y directions have the same lattice constant, and have identical symmetry. Then the planes have the same crystal symmetry as a square. In polar coordinates the angle θ gives

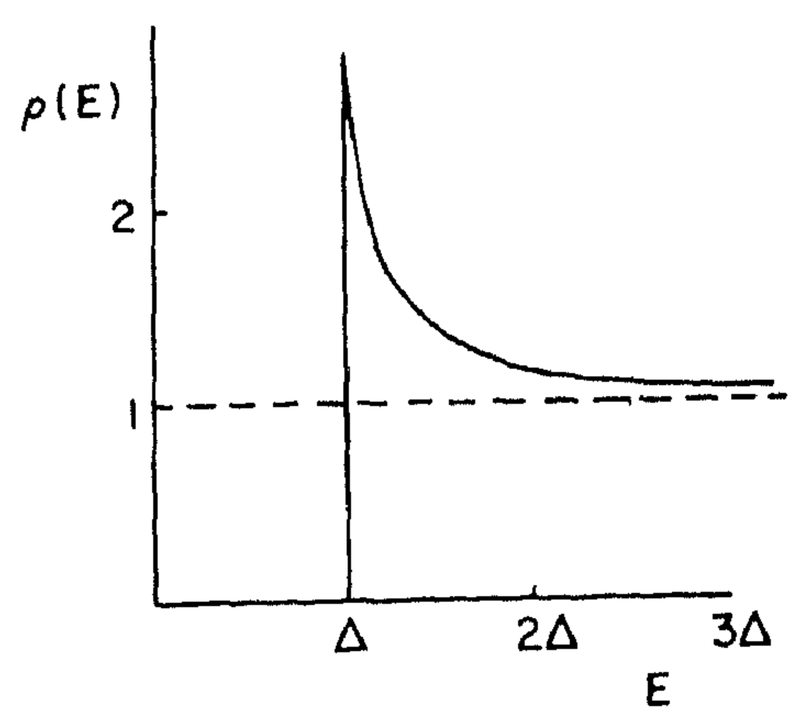


FIGURE 10.7 BCS density of states of a superconductor.

the wave vector direction in the plane ($\tan\theta = k_y/k_x$). In this case the superconducting order parameter for d-wave symmetry is either of the two functions

$$\Delta_{d1}(\theta) = \Delta_1 \cos(2\theta) \tag{10.79}$$

$$\Delta_{d2}(\theta) = \Delta_2 \sin(2\theta) \tag{10.80}$$

Similarly, another possible order parameter called "anistropic s-wave" is

$$\Delta_s(\theta) = \Delta_0 + \Delta_a \cos(4\theta) \tag{10.81}$$

Note that the term $\cos(4\theta)$ preserves the symmetry of the crystal: at $\theta = \pi/2$ the function has the same value as at $\theta = 0$. The x- and y-axes are identical for anisotropic s-wave superconductivity. In contrast, the d-wave superconductivity breaks crystal symmetry.

An important feature of d-wave superconductivity is that the order parameter changes sign as the angle θ is varied. There are nodes, where the gap function vanishes. The anisotropic s-wave can also have nodes if $\Delta_a > \Delta_0$, but that is unlikely.

10.2. SUPERCONDUCTING TUNNELING

The most important verifications of the BCS theory came from electron tunneling experiments. They measured the energy gap, as a function of temperature, in perfect agreement with the BCS theory. They also measured the density of states function $\rho(E)$. Later, Josephson (1962) predicted the coherent tunneling of pairs, which was also quickly observed. These experiments provided a detailed verification of the BCS theory.

The theory of electron tunneling between two metal electrodes, through a barrier consisting of a metal oxide, was derived in Sec. 8.6. The tunneling Hamiltonian was used to obtain the following expression for the tunneling current

$$I = I_S + I_J \tag{10.82}$$

$$I_S = ie[U_{\text{ret}}(-eV) - U_{\text{adv}}(-eV)] \tag{10.83}$$

$$= -2e\,\text{Im}[U_{\text{ret}}(-eV)] \tag{10.84}$$

$$U(i\omega) = -\int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau A(\tau) A^\dagger(0)\rangle \tag{10.85}$$

$$A = \sum_{\mathbf{kp}\sigma} T_{\mathbf{kp}} C^\dagger_{\mathbf{k}\sigma} C_{\mathbf{p}\sigma} \tag{10.86}$$

The term I_S is the single-particle tunneling. The term I_J is for the Josephson effect, and is evaluated below. In Sec. 8.6 this expression was evaluated, and reduced to an expression in terms of spectral functions and occupation numbers

$$I_S = 2e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \int \frac{d\xi}{2\pi} A_L(\mathbf{p}, \xi + eV) A_R(\mathbf{k}, \xi)[n_F(\xi) - n_F(\xi + eV)] \tag{10.87}$$

The subscripts L, R refer to which electrode. The tunneling matrix element $T_{\mathbf{kp}}$ can be approximated as a constant for the small voltages used in superconductive tunneling. Then

the summations over wave vectors give the interacting density of states $N_{L,R}(\varepsilon)$ on the two sides of the junction.

$$I_S = 4\pi e|T|^2 \int d\varepsilon N_R(\varepsilon) N_L(\varepsilon + eV)[n_F(\varepsilon) - n_F(\varepsilon + eV)]$$

This expression is also valid in a superconductor with an appropriate interpretation of the density of states.

10.2.1. Normal Superconductor

Electron tunneling experiments are often done with different metals on the two sides of the junction. Since different metals have different transition temperatures (and some are not superconducting), it is quite easy to arrange that one metal be normal and the other be a superconductor. This experiment is interesting, to the theorist, because it provides a direct measurement of the BCS density of states $\rho(E)$. Take the right side to be superconducting and the left side to be normal. The derivation is simple, since the many-body expression (10.87) is valid in this case. Just use the respective spectral functions for the normal and superconducting sides of the tunneling junction:

$$A_R(\mathbf{k}, \varepsilon) = 2\pi[u_k^2 \delta(\varepsilon - E_k) + v_k^2 \delta(\varepsilon + E_k)] \tag{10.88}$$

$$A_L(\mathbf{p}, \varepsilon + eV) = 2\pi\delta(\varepsilon + eV - \xi_p) \tag{10.89}$$

$$\begin{aligned} I_S = 4\pi e|T|^2 \sum_{\mathbf{kp}} \{ & u_k^2 \delta(eV + E_k - \xi_p)[n_F(E_k) - n_F(\xi_p)] \\ & + v_k^2 \delta(eV - E_k - \xi_p)[n_F(-E_k) - n_F(\xi_p)]\} \end{aligned} \tag{10.90}$$

Next use the fact that $n_F(-E) = 1 - n_F(E)$. Consider the situation at zero temperature, where the density of excitations in the superconductor is negligible $[n_F(E) \approx (e^{\beta E} + 1)^{-1} \approx e^{-\beta E} \approx 0]$. The variable change (8.337) brings us to

$$\begin{aligned} I_s = 4\pi e|T|^2 N_R N_L \int_{-\infty}^{\infty} d\xi_p \int_{-\infty}^{\infty} d\xi_k \{ & v_k^2 \delta(eV - E_k - \xi_p)[1 - n_F(\xi_p)] \\ & - u_k^2 \delta(eV + E_k - \xi_p) n_F(\xi_p)\} \end{aligned} \tag{10.91}$$

The limits of integration have been extended to infinity, which has no effect on the result since the limits are narrowly confined. There are two terms in the curly braces: the first is positive, and the second is negative. The first will contribute only when $eV > 0$ (actually $eV > \Delta$), and I_S is positive in this case because this direction of electron current is defined to be positive. The second term contributes only when $eV < -\Delta$, and I_S is negative in this case, in agreement with the definition, since I_S has the same sign as V.

At zero temperature, the first term has a factor $1 - n_F(\xi_p)$ which limits $\xi_p > 0$. The integral $d\xi_k$ extends over positive and negative values of ξ_k. Factors which depend on

$E_k = \sqrt{\xi_k^2 + \Delta^2}$ are the same for both signs of ξ_k. The term linear in ξ_k in the coherence factor v_k^2 in (10.63) averages to zero because of the cancellation of the $+\xi_k$ and $-\xi_k$ parts:

$$\int_{-\infty}^{\infty} d\xi_k v_k^2 f(E_k) = \frac{1}{2}\int_{-\infty}^{\infty} d\xi_k \left(1 - \frac{\xi_k}{E_k}\right) f(E_k) = \frac{1}{2}\int_{-\infty}^{\infty} d\xi_k f(E_k)$$
$$= \int_0^{\infty} d\xi_k f(E_k) = \int_{\Delta}^{\infty} dE_k \rho(E_k) f(E_k) \tag{10.92}$$

where f is any function of E_k. The integration is changed to dE_k, which produces a factor of the density of states $\rho(E) = d\xi/dE$ in the superconductor. The integral $\int d\xi_p$ can be used to eliminate the delta function for energy conservation $\xi_p = eV - E_k > 0$. Since $\xi_p > 0$ then $eV > E_k \geq \Delta$, which gives for $eV > 0$

$$I_{SN} = \frac{\sigma_0}{e}\Theta(eV - \Delta)\int_{\Delta}^{eV} dE\rho(E) = \frac{\sigma_0}{e}\Theta(eV - \Delta)\int_{\Delta}^{eV} \frac{EdE}{\sqrt{E^2 - \Delta^2}}$$
$$= \frac{\sigma_0}{e}\Theta(eV - \Delta)\sqrt{(eV)^2 - \Delta^2} \tag{10.93}$$

The current is a simple square root function of the voltage. This type of dependence is observed experimentally.

Another quantity which can be measured (by lock-in amplifier techniques) is the dynamical conductance dI/dV as a function of voltage:

$$\left(\frac{dI_S}{dV}\right)_{SN} = eV\sigma_0 \frac{\Theta(eV - \Delta)}{\sqrt{(eV)^2 - \Delta^2}} \tag{10.94}$$

$$\frac{\left(\frac{dI_S}{dV}\right)_{SN}}{\left(\frac{dI_S}{dV}\right)_{NN}} = \Theta(eV - \Delta)\frac{eV}{\sqrt{(eV)^2 - \Delta^2}} = \rho(eV) \tag{10.95}$$

The experimental technique is to compare dI/dV for the tunneling between the normal metal and the superconductor $(dI/dV)_{SN}$, with that for the tunneling between both metals when they are normal $(dI/dV)_{NN} = \sigma_0$. Low-temperature superconductors can be made normal by the application of a small magnetic field, so that the two measurements are done at the same low temperature. The ratio of these two experimental quantities provides a direct measurement of the BCS density of states $\rho(eV)$. The current is antisymmetric in voltage, which can be shown from the second term in (10.91).

Figure 10.8 shows the experimental results of Giaever *et al.* (1962) for dI/dV of a tunnel junction between Pb and Mg. The experiments were done at a temperature of 0.3 K with a ^{3}He refrigerator. At this low temperature, the Mg is normal, and the Pb is well below its transition temperature of $T_c = 7.2$ K. The factor $n_F(E) \sim \exp(-\beta\Delta)$ is indeed small, and there should be few thermal excitations in the superconducting Pb. The figure shows the derivative (dI/dV), which well illustrates the BCS density of states. Another feature of their data are bumps in dI/dV at higher voltages. These are due to phonons and are not predicted by the BCS theory. They are explained by the strong coupling theory. These experimental results by Giaever *et al.* (1962) were one of the first experimental indications of the need for strong coupling theory.

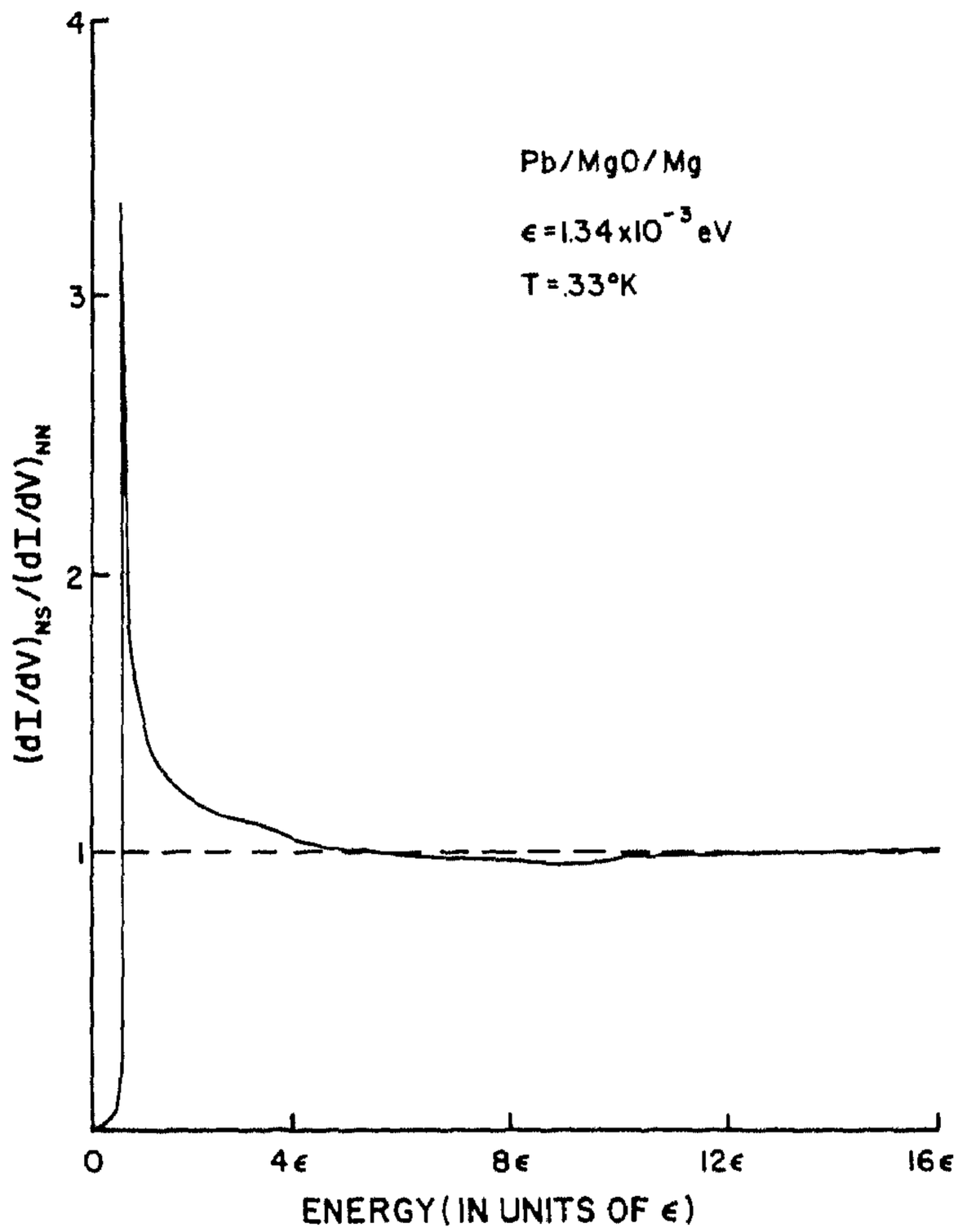


FIGURE 10.8 Electron tunneling between a normal metal (Mg) and a superconductor (Pb). The relative conductance of a Pb–MgO–Mg sandwich plotted against energy. At higher energies there are definite divergences from the BCS density of states, as can be seen from the bumps in the experimental curve. *Source:* Giaever *et al.* (1962) (used with permission).

The measured current is an antisymmetric function of the voltage $I(-V) = -I(V)$ for the small voltages of interest in this experiment. An energy gap of $\Delta = 1.34$ meV is observed in the tunneling current at zero temperature (which can only be approached asymptotically), and no current flows unless $|eV| > \Delta$. An interpretation of these results is provided in Fig. 10.9. Figure 10.9(a) shows the two sides of the tunnel junction with no applied voltage $V = 0$. The chemical potentials are the same on the two sides of the junction. On the normal side (left) the chemical potential divides the occupied states from the empty states in the Fermi sea of electrons. On the superconductor side (right), the chemical potential is the energy of the paired electrons. The excitation spectrum is shown as the hyperbolic-shaped line above the chemical potential. No net current flows when $V = 0$.

The situation for a forward voltage $eV > \Delta$ is shown in Fig. 10.9(b). The electrons from the normal side of the junction can tunnel through and become a quasiparticle excitation of energy E_k on the superconducting side. The arrow, which shows the tunneling path, is drawn horizontal to indicate the energy-conserving aspect of tunneling. The electron cannot go into the pair state on the right as a single step in the tunneling, since it needs two electrons for this. It must become an excited quasiparticle during its tunneling step. Later it will find some other excited electron and join with it to become a pair state. This pairing is expected to happen well after the tunneling process and so will not affect the tunneling rate.

The situation for reverse bias is shown in Fig. 10.9(c). At zero temperature, all the electrons in the superconductor are in the pair states at the chemical potential, and none are thermally excited to the excitation state. The single-particle tunneling process must break up the pair state. For each electron which tunnels, its partner in the initial pair state must become

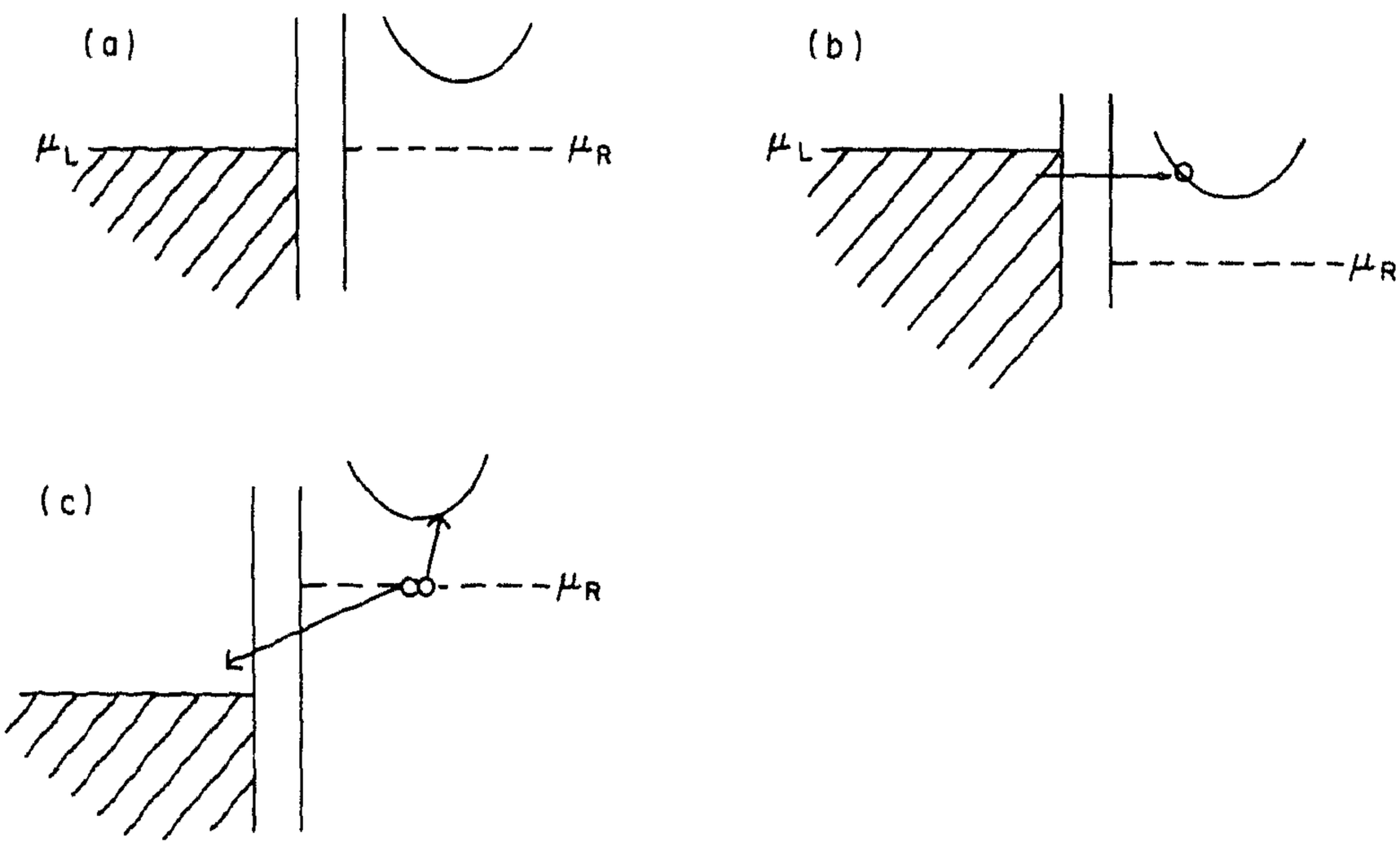


FIGURE 10.9 Electron tunneling between a normal metal and a superconductor at $T = 0$. (a) No applied voltage. (b) For $eV > \Delta$, electrons can tunnel from the normal metal and becomes a quasiparticle in the superconductor. (c) For $eV < -\Delta$, a pair is broken in the superconductor: one electron tunnels, while the other becomes a quasiparticle in the superconductor.

an excited quasiparticle. The electron which tunnels does not do this at constant energy, since it must give up some of its energy to its partner to permit it to overcome the excitation gap Δ. Energy conservation can be understood as follows. The initial energy of the paired electrons is $E_{\text{initial}} = 2\mu_R$. The final energy is $\mu_R + E_k$ for the one in the superconductor and $\mu_L + \xi_p$ for the one in the normal metal. The total energy of the original pair in the final state is

$$E_{\text{final}} = \mu_R + \mu_L + \xi_p + E_k = E_{\text{initial}} = 2\mu_R \tag{10.96}$$

The entire process conserves energy, so this is set equal to the initial energy. The quasiparticle energy in the normal metal ξ_p must be positive since the electron must go into an unoccupied state. The condition for the tunneling current to flow is:

$$\mu_R - \mu_L = eV = \xi_p + E_k > \Delta \tag{10.97}$$

This condition is that the magnitude of the voltage must be larger than Δ, which is the same condition for the other direction of bias voltage. The actual physical processes for the two directions of voltage are dissimilar, since in one direction there is direct pair breaking.

10.2.2. Two Superconductors

The electron tunneling between two superconductors has two kinds of currents. One is the tunneling of a single electron, which is similar to that between normal metals or between normal metals and superconductors. The other is the tunneling of pairs, which is the Josephson effect. This interesting subject will be described in the next section. Now consider only the single-particle tunneling, which can happen between two similar or two dissimilar superconductors. Take the case where they are different, so there are gap functions Δ_R and Δ_L for the two sides of the junction. At any point in the calculation the two superconductors can be made identical, theoretically, by just setting $\Delta_R = \Delta_L$.

The many-body formula (8.344) is still valid when the two sides are superconductors

$$I = 2e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} A(\mathbf{k}, \varepsilon) A(\mathbf{p}, \varepsilon + eV)[n_F(\varepsilon) - n_F(\varepsilon - eV)] \tag{10.98}$$

The spectral function for the superconducting BCS Green's function is used for both sides of the tunneling junction:

$$A(\mathbf{k}, \varepsilon) = 2\pi[u_k^2 \delta(\varepsilon - E_k) + v_k^2 \delta(\varepsilon + E_k)] \tag{10.99}$$

$$A(\mathbf{p}, \varepsilon + eV) = 2\pi[u_p^2 \delta(\varepsilon + eV - E_p) + v_p^2 \delta(\varepsilon + eV + E_p)] \tag{10.100}$$

Their product gives four terms, which are regrouped using the identity that $n_F(-E) = 1 - n_F(E)$:

$$\begin{aligned} I = 4\pi e \sum_{\mathbf{kP}} |T_{\mathbf{kp}}|^2 \{ & u_k^2 u_p^2 \delta(eV + E_k - E_p)[n_F(E_k) - n_F(E_p)] \\ & + u_k^2 v_p^2 \delta(eV + E_k + E_p)[n_F(E_k) - n_F(-E_p)] \\ & + v_k^2 u_p^2 \delta(eV - E_k - E_p)[n_F(-E_k) - n_F(E_p)] \\ & + v_k^2 v_p^2 \delta(eV - E_k + E_p)[n_F(-E_k) - n_F(-E_p)]\} \end{aligned} \tag{10.101}$$

$$\begin{aligned} I = 4\pi e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \{ & [1 - n_F(E_p) - n_F(E_k)][v_k^2 u_p^2 \delta(eV - E_p - E_k) \\ & - u_k^2 v_p^2 \delta(eV + E_k + E_P)] + [n_F(E_k) - n_F(E_p)] \\ & \times [u_k^2 u_p^2 \delta(eV + E_k - E_p) - v_k^2 v_p^2 \delta(eV - E_k + E_p)]\} \end{aligned} \tag{10.102}$$

This formula is the most general result, which is valid at nonzero temperatures. The variable change to $d\xi_k d\xi_p$ is done next. These integrals over the coherence factors u^2 and v^2 will eliminate their linear term in ξ, as described in (10.92). Then the variables can be changed to $E = E_k = \sqrt{\xi_k^2 + \Delta_R^2}$ and $E' = E_p = \sqrt{\xi_p^2 + \Delta_L^2}$, which brings us to the expression

$$\begin{aligned} I = 4\pi e N_L N_R |T|^2 & \int_{\Delta_R}^{\infty} dE \rho_R(E) \int_{\Delta_L}^{\infty} dE' \rho_L(E') \\ & \times \{[1 - n_F(E) - n_F(E')][\delta(eV - E - E') - \delta(eV + E + E')] \\ & + [n_F(E) - n_F(E')][\delta(eV + E - E') - \delta(eV - E + E')]\} \end{aligned} \tag{10.103}$$

Some of the terms are positive, while others are negative. They will ensure that I is positive for $V > 0$ and negative for $V < 0$. These integrals must be evaluated numerically at nonzero temperatures. The physics is illustrated by taking the simplest case, which is zero temperature and identical superconductors $\Delta_L = \Delta_R$. At zero temperature, set to zero all the thermal factors $n_F(E)$ and $n_F(E')$:

$$\begin{aligned} I &= \frac{\sigma_0}{e} \int_{\Delta}^{\infty} dE \rho(E) \int_{\Delta}^{\infty} dE' \rho(E')[\delta(eV - E - E') - \delta(eV + E + E')] \\ \sigma_0 &= 4\pi e^2 N_R N_L |T|^2 \end{aligned} \tag{10.104}$$

The two terms in square brackets give identical results, where the first is for $eV > 0$ and the second for $eV < 0$, in such a way that $I(-V) = -I(V)$. Only the first term needs to be done, where $E' = eV - E$ and $eV > 2\Delta$:

$$I = \left(\frac{\sigma_0}{e}\right)\Theta(eV - 2\Delta)\int_{\Delta}^{eV-\Delta} dE\rho_R(E)\rho(eV - E)$$

$$= \left(\frac{\sigma_0}{e}\right)\Theta(eV - 2\Delta)\int_{\Delta}^{eV-\Delta} dE\frac{E(eV - E)}{\sqrt{E^2 - \Delta^2}\sqrt{(eV - E)^2 - \Delta^2}} \tag{10.105}$$

This integral is done in the following way. Change variables of integration to x, where $2E = eV + x(eV - 2\Delta)$. After some algebra, the integral is in the form of complete elliptic integrals:

$$\alpha = \frac{eV - 2\Delta}{eV + 2\Delta} \tag{10.106}$$

$$I = \left(\frac{\sigma_0}{e}\right)\frac{\Theta(eV - 2\Delta)}{eV + 2\Delta}\int_{-1}^{1} dx\frac{(eV)^2 - x^2(eV - 2\Delta)^2}{[(1 - x^2)(1 - \alpha^2x^2)]^{1/2}}$$

$$= \left(\frac{\sigma_0}{e}\right)\Theta(eV - 2\Delta)\left\{\frac{(eV)^2}{eV + 2\Delta}K(\alpha) - (eV + 2\Delta)[K(\alpha) - E(\alpha)]\right\} \tag{10.107}$$

Equation (10.107) is the final result, which is compared with the experiments. One check on the correctness of the answer is to set the energy gap $\Delta = 0$. The coefficient of $K(\alpha)$ vanishes, so the part in curly braces becomes $eVE(1) = eV$. The current is $I = \sigma_0 V$, which is the correct result, since in the absence of superconductivity the tunneling device is a resistor at low voltage with the conductance σ_0.

The tunneling current is zero for voltages less than twice the gap function 2Δ. It is interesting to examine the threshold phenomenon, for voltages slightly larger than 2Δ. Since $K(0) = E(0) = \pi/2$, at $eV = 2\Delta^+$ the current is

$$I(2\Delta^+) = \left(\frac{\sigma_0}{e}\right)\frac{\pi\Delta}{2} = \sigma_0 V\frac{\pi}{4} \tag{10.108}$$

It has a nonzero value, so there is a discontinuous jump in the tunneling current at $eV = 2\Delta$. The current above threshold is $\pi/4$ of the value of a normal tunneling current at the same voltage. Exactly this behavior was found by Giaever *et al.* (1962). Their data for a Sn–SnO_x–Sn junction is shown in Fig. 10.10. The entire experimental curve fits the theoretical curve from (10.107), except for the finite slope on the "discontinuity." They also point out an interesting feature of this discontinuous jump in current: the discontinuity does not get thermally broadened with increasing temperature. As the temperature is increased, the energy gaps get smaller, so the value 2Δ of the jump moves to small voltages. However, the BCS energy gap does not get smeared at higher temperature; it only gets smaller. At higher temperatures, the nonzero value of $n_F(E)$ will lead to some tunneling current for $eV < 2\Delta$, but there is still a discontinuous jump at this critical voltage.

The physical process of the single-particle tunneling is shown in Fig. 10.11. For electron flow to the right, a pair on the left is broken up, and one electron tunnels, while the other becomes a quasiparticle on the left. The initial energy of the pair is $E_i = 2\mu_L$, and the final energy of the pair is $E_f = (\mu_L + E_L) + (\mu_R + E_R)$. By equating these two energies, we derive

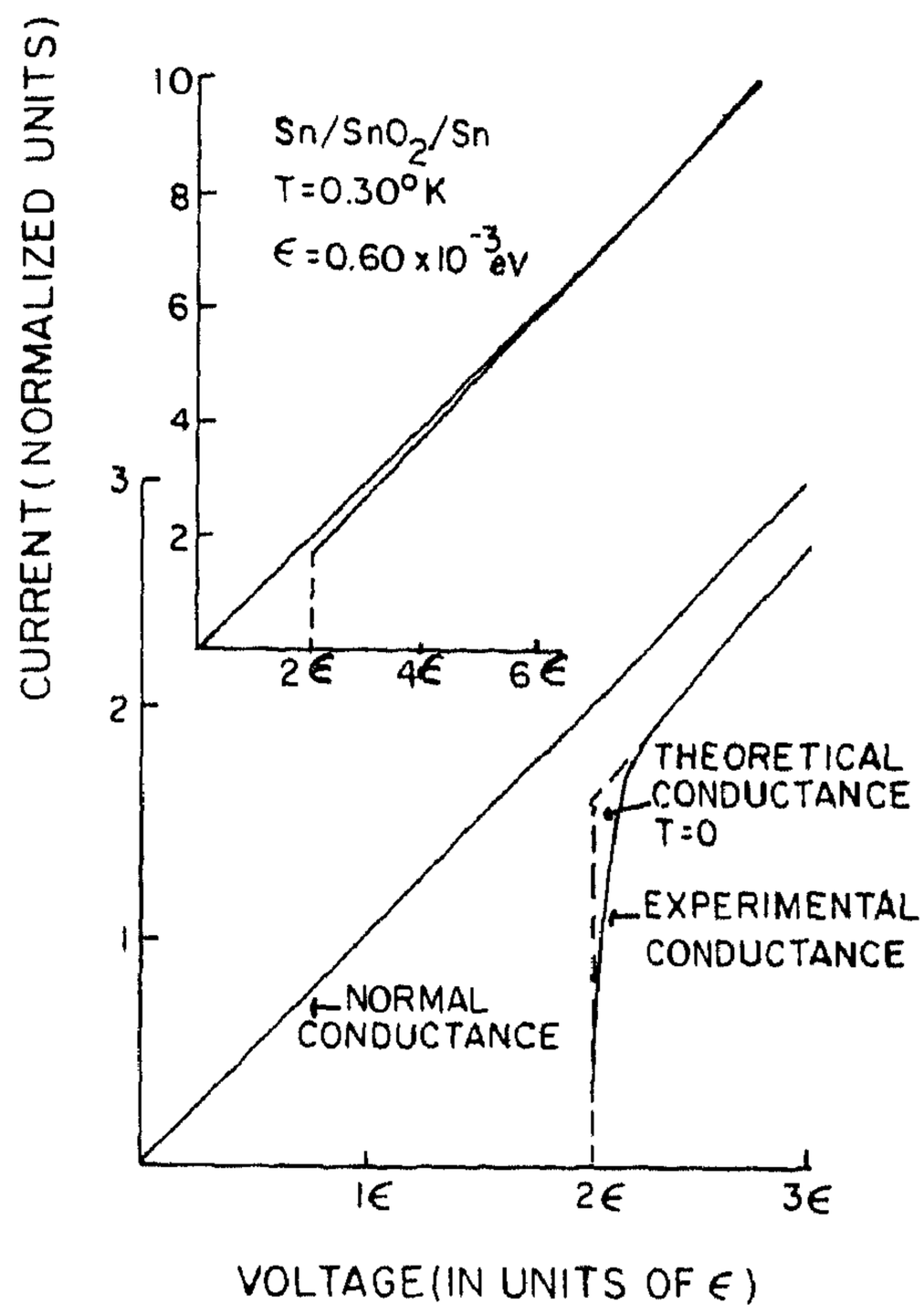


FIGURE 10.10 Electron tunneling between two identical superconductors. Current voltage characteristic for a sandwich compared with the BCS theory. The experimental curve has a finite slope at 2Δ because of the anisotropic nature of the energy gap. *Source:* Giaever *et al.* (1962) (used with permission).

$\mu_L - \mu_R = eV = E_L + E_R$. The summation of the two excitation energies must be equal to the applied voltage, which is in agreement with the theoretical formulas.

The reader should now be convinced that the BCS theory rather well describes the theory of electron tunneling in low temperature superconductors. Indeed, the BCS theory describes rather well all the properties of weakly coupled superconductors, and other good agreements between theory and experiment will be shown later. Several severe approximations were made in the derivation of the BCS theory. Apparently none of these really matter. Two of the major approximations were (1) calling the gap function $\Delta(p)$ a constant Δ and (2) replacing the retarded (i.e., frequency–dependent) interaction between electrons by a static potential, which is a constant. A constant potential in Fourier space is a delta function in real space, so it was assumed that the interaction between electrons is localized in space. The electrons themselves

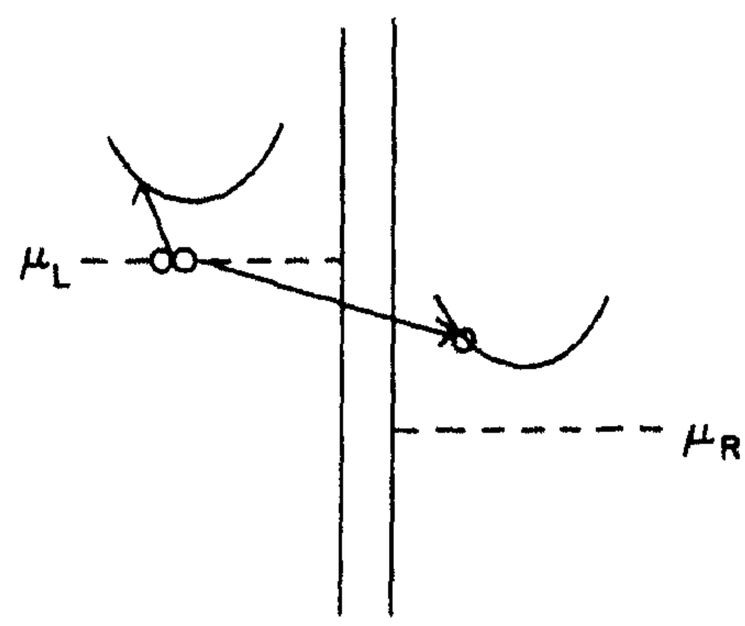


FIGURE 10.11 Physical process of electron tunneling between two superconductors. A pair breaks on the left, one electron tunnels, and both electrons become excited quasiparticles.

are not localized. Indeed, the pairing orbits are extended in space, but the particles interact only when they are in contact. In any case, the essential feature of BCS theory is the existence of the additional correlation functions $\mathscr{F}$ and $\mathscr{F}^\dagger$. This feature must be right, and the particular method of getting to these correlation functions is not very crucial.

Electron tunneling is a less successful experimental tool for the investigation of high temperature superconductivity. Although many experiments have been done, the results are inconsistent and difficult to interpret. There are several problems. The cuprates are layered materials. The electron conduction is within a layer, and the layers are weakly coupled. That means the top few layers of a material may not even be superconducting. Tunneling down to the top layer may be going into a normal layer, even if the bulk is superconducting. The other geometry is to tunnel into the layers from the edge. In this case, the electron is injected into a single layer, which is a two-dimensional electron gas. The last chapter discussed MND effects, which is the renormalization of a transition intensity due to the process of injecting an electron into an electron gas. Such effects are accentuated in two dimensions. So the spectral function measured in tunneling may be severely distorted due to electron–electron interactions (Shekhtman and Glazman, 1995). Finally, if the gap has d-wave symmetry there are nodes where the gap vanishes. Depending upon the directional dependence of the tunneling, there may be no energy gap to observe.

10.2.3. Josephson Tunneling

The derivation of the tunneling current led to two terms, which are I_S and I_J. They are the single-particle term and the Josephson term. The single-particle tunneling has been discussed in great detail, and now examine the Josephson effect. The discussion of this current contribution closely follows Ambegaokar and Baratoff (1962, 1963). The Josephson current was derived in Sec. 8.6

$$I_J(t) = e\int_{-\infty}^{\infty} dt'\Theta(t-t')\{e^{-ieV(t+t')}\}\langle[A(t), A(t')]\rangle - e^{ieV(t+t')}\langle[A^\dagger(t), A^\dagger(t']\} \tag{10.109}$$

The single most interesting feature of this equation is that it contains, in the two exponents, the sum of the two times $t + t'$. This dependence has not been previously encountered in any of the correlation functions, since they usually come out containing $t - t'$. The factor of $t + t'$ is the first indication, mathematically, that the Josephson effect is rather special. The problem of $t + t'$ is solved by writing it as $2t + t' - t$ and pulling the factor $2t$ outside of the correlation function. The current is written as

$$I_J(t) = e\left\{e^{-2iteV/\hbar}\int_{-\infty}^{\infty} dt'\Theta(t-t')e^{ieV(t-t')}\langle[A(t), A(t')]\rangle + \text{h.c.}\right\}$$

Define the retarded correlation function

$$\Phi_{\text{ret}}(t-t') = -i\Theta(t-t')\langle[A(t), A(t')]\rangle \tag{10.110}$$

$$\Phi_{\text{ret}}(eV) = \int_{-\infty}^{\infty} dt e^{ietV(t-t')}\Phi_{\text{ret}}(t-t') \tag{10.111}$$

$$I_J(t) = -ie\left[e^{-2ietV/\hbar}\Phi_{\text{ret}}(eV) - e^{2ietV/\hbar}\Phi_{\text{adv}}(eV)\right] \tag{10.112}$$

The part in brackets is purely imaginary since it is a quantity minus its complex conjugate. The Josephson current is

$$I_J(t) = 2e\,\mathrm{Im}[e^{-2ietV/\hbar}\Phi_{\mathrm{ret}}(eV)] \tag{10.113}$$

It is important to include the factor $\exp(-2iteV/\hbar)$ in the expression whose imaginary part is evaluated. The Josephson tunneling is not just the imaginary part of a retarded function—it is not just a spectral function. Instead, one must multiply by the phase factor $\exp(-2iteV/\hbar)$ before taking the imaginary part. The phase factor contains most of the interesting physics. For a nonzero voltage, the current is actually oscillatory in time, with a frequency $\omega = 2eV/\hbar$. This prediction of the theory is so absolute, and simple, that the frequency can be used as a method of measuring the ratio of fundamental constants e/h (see Parker *et al.*, 1969).

The first task is to develop a method of calculating the retarded correlation function $\Phi_{\mathrm{ret}}(eV)$. As usual, start from a Matsubara function,

$$\begin{aligned}\Phi(i\omega) &= -\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau A(\tau)A(0)\rangle \\ &= -\sum_{\mathbf{kp}\sigma}\sum_{\mathbf{k}'\mathbf{p}'\sigma'} T_{\mathbf{kp}}T_{\mathbf{k}'\mathbf{p}'}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau C^\dagger_{\mathbf{k}\sigma}(\tau)C_{\mathbf{p}\sigma}(\tau)C^\dagger_{\mathbf{k}'\sigma'}C_{\mathbf{p}'\sigma'}\rangle \end{aligned}\tag{10.114}$$

and later make the analytical continuation $i\omega \to eV + i\delta$. The spin index has been explicitly added to this expression. The correlation function can be factored into a term for the left and right sides of the tunnel junction. In this case the factoring produces the Green's function $\mathscr{F}$ and $\mathscr{F}^\dagger$:

$$\begin{aligned}\Phi(i\omega) &= 2\sum_{\mathbf{kp}} T_{\mathbf{kp}}T_{-\mathbf{k},-\mathbf{p}}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau C^\dagger_{\mathbf{k}\uparrow}(\tau)C^\dagger_{-\mathbf{k}\downarrow}(0)\rangle\langle T_\tau C_{-\mathbf{p}\downarrow}(0)C_{\mathbf{p}\uparrow}(\tau)\rangle \\ &= 2\sum_{\mathbf{kp}} T_{\mathbf{kp}}T_{-\mathbf{k},-\mathbf{p}}\int_0^\beta d\tau e^{i\omega_n\tau}\mathscr{F}^\dagger(\mathbf{k},\tau)\mathscr{F}(\mathbf{p},-\tau) \qquad (10.115)\\ &= 2\sum_{\mathbf{kp}} T_{\mathbf{kp}}T_{-\mathbf{k},-\mathbf{p}}\frac{1}{\beta}\sum_{ip}\mathscr{F}^\dagger(\mathbf{k},ip)\mathscr{F}(\mathbf{p},ip-i\omega) \qquad (10.116)\end{aligned}$$

The factor of 2 is for spin arrangement. One spin configuration is shown explicitly, and the other one gives the same result. The Feynman diagram for this tunneling is shown in Fig. 10.12. There is an $\mathscr{F}^\dagger$ on the right side of the junction and $\mathscr{F}$ on the left, where the arrow conventions agree with those set forth in Fig. 10.5. The direction of the arrows is continuous through the symbol T in a circle, which must always happen with any diagram where particle number is conserved.

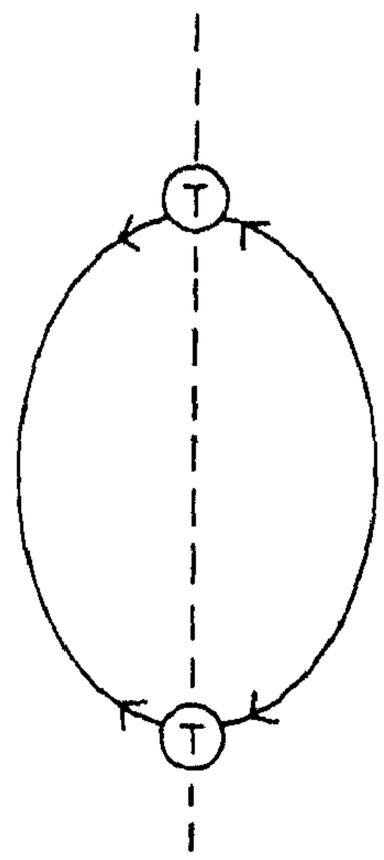


FIGURE 10.12 Feynman diagram for Josephson tunneling.

The summation over Matsubara frequencies in (10.116) is done by the usual method of contour integral:

$$\frac{1}{\beta}\sum_{ip}\mathscr{F}^{\dagger}(\mathbf{k}, ip)\mathscr{F}(\mathbf{p}, ip-\omega) \tag{10.117}$$

$$= \frac{\Delta_L\Delta_R}{3E_kE_p}\left\{[1-n_F(E_f)-n_F(E_p)]\left(\frac{1}{i\omega+E_p+E_k}-\frac{1}{i\omega-E_k-E_p}\right)\right.$$

$$\left.+[n_F(E_p)-n_F(E_k)]\left(\frac{1}{i\omega+E_p-E_k}-\frac{1}{i\omega+E_k-E_p}\right)\right\}$$

$$\Phi(i\omega) = \Delta_L\Delta_R\sum_{\mathbf{kp}}\frac{T_{\mathbf{kp}}T_{-\mathbf{k},-\mathbf{p}}}{2E_kE_p} \tag{10.118}$$

$$\times\left\{[1-n_F(E_f)-n_F(E_p)]\left(\frac{1}{i\omega+E_p+E_k)}-\frac{1}{i\omega-E_k-E_p}\right)\right.$$

$$\left.+[n_F(E_p)-n_F(E_k)]\left(\frac{1}{i\omega+E_p-E_k}-\frac{1}{i\omega+E_k-E_p}\right)\right\}$$

This expression is rather complicated, particularly for nonzero temperatures and different superconductors. The physics may be best understood by examining the simplest situation, which is zero temperature $[n_F(E)=0]$ and identical superconductors $(\Delta_L=\Delta_R)$. With the analytical continuation $i\omega\to eV+i\delta$, consider the quantity

$$\Phi_{\text{ret}}(eV) = \Delta^2\sum_{\mathbf{kp}}\frac{T_{\mathbf{kp}}T_{-\mathbf{k},-\mathbf{p}}}{2E_kE_p}\left(\frac{1}{eV+E_p+E_k+i\delta}-\frac{1}{eV-E_k-E_p+i\delta}\right) \tag{10.119}$$

An important feature of these equations is that the tunneling matrix elements $T_{\mathbf{kp}}$ are not necessarily real, so the combination $T_{\mathbf{kp}}T_{-\mathbf{k},-\mathbf{p}}$ is not real. Write this factor as an amplitude

$|T|^2$ and a phase factor $\exp(i\Phi)$ and consider both of these to be independent of **k** and **p**. The summations over **k** and **p** can be changed to integrations over $d\xi_k d\xi_p$ and then to $dE_k dE_p$:

$$\Phi_{\text{ret}}(eV) = \frac{1}{2e} J_S(eV) e^{i\Phi} \tag{10.120}$$

$$J_S(eV) = 4e\Delta^2 |T|^2 N_L N_R \int_\Delta^\infty dE \frac{\rho(E)}{E} \int_\Delta^\infty dE' \frac{\rho(E')}{E'} \left(\frac{1}{eV + E + E'} - \frac{1}{eV - E - E'} \right) \tag{10.121}$$

$$= \Delta \left(\frac{\sigma_0}{e} \right) K \left(\frac{eV}{2\Delta} \right) \tag{10.122}$$

The integral is an elliptic integral. The retarded correlation function is written as a phase factor times a function of voltage $J_S(eV)$, where J_S is real for $|eV| < 2\Delta$. The Josephson current can be written as the imaginary part of (10.113), or

$$I_S(t) = J_S(eV) \sin(\omega t + \Phi) \tag{10.123}$$

$$\omega = \frac{2eV}{\hbar} \tag{10.124}$$

The Josephson current oscillates with time, with a frequency $\omega = 2eV/\hbar$. The amplitude of the oscillation is determined by the function of voltage $J_S(eV)$, where it is assumed $|eV| < 2\Delta$. The easiest way to evaluate $J_S(eV)$ at nonzero temperatures is to use a different sequence of steps. Ambegaokar and Baratoff showed that it is best to return to (10.116) and to do the two integrals $d\xi_k d\xi_p$ before the summation over Matsubara frequency ip_n. This step uses the result

$$\int_{-\infty}^{\infty} d\xi_k \mathscr{F}^\dagger(k, ip) = \Delta \int_{-\infty}^{\infty} \frac{d\xi_k}{\xi_k'^2 + \Delta^2 + p_n^2} = \frac{\pi\Delta}{\sqrt{\Delta^2 + p_n^2}} \tag{10.125}$$

The function $J_S(eV)$ is the retarded part ($i\omega \to eV$) of the function $J_S(i\omega)$:

$$J_S(i\omega) = 4e \sum_{\mathbf{kp}} |T_{\mathbf{kp}}|^2 \frac{1}{\beta} \sum_{ip} \mathscr{F}^\dagger(\mathbf{k}, ip) \mathscr{F}(\mathbf{p}, ip - i\omega) \tag{10.126}$$

$$= 4\pi^2 e N_R N_L |T|^2 \frac{1}{\beta} \sum_{ip} \frac{\Delta_R \Delta_L}{\sqrt{\Delta_R^2 + p_n^2} \sqrt{\Delta_L^2 + (p_n - \omega_m)^2}} \tag{10.127}$$

The summation over Matsubara frequencies ip_n should be done before the step of analytical continuation $i\omega \to eV$. An error is made if these two steps are reversed, except at the point $eV = 0$. At zero voltage, the result for identical superconductors is [see (10.67)]

$$J_S(eV = 0) = 4e\pi^2 N^2 \Delta^2 |T|^2 \frac{1}{\beta} \sum_{ip_n} \frac{1}{\Delta^2 + p_n^2} \tag{10.128}$$

$$= 2e\pi^2 N^2 \Delta |T|^2 \tanh\left(\frac{\beta\Delta}{2} \right) \tag{1.129}$$

At zero temperature there is the particularly simple result that

$$J_S(0) = \left(\frac{\sigma_0}{e} \right) \left(\frac{\pi\Delta}{2} \right) \tag{10.130}$$

The coefficient of the Josephson current is the same current found in a normal junction at a voltage of $eV = \pi\Delta/2$.

Josephson tunneling has a wide variety of behavior in junctions in static electric and magnetic fields. A very simple characteristic current is obtained in the case where there is neither a voltage $V = 0$ nor a magnetic field. Then one gets that the Josephson current is

$$I_J = J_S(0)\sin(\Phi) \tag{10.131}$$

where Φ is the phase difference between the two superconductors on the two sides of the tunnel junction.

The experimental arrangement is to put a current I through the tunnel junction, so that current flows through the interface oxide. The voltage change V is measured as a function of the applied current I. It is found that a dc current can flow without any voltage V. This behavior is sketched in Fig. 10.13. The dc current is the coherent tunneling of electron pairs from one superconductor to another. The pair tunneling costs no energy when the two superconductors have the same chemical potential. When the experimentalist increases the current, the phase difference Φ between the two superconductors adjusts to make the Josephson current $J_S(0)\sin\Phi$ just the experimental value I of current. The phase increase continues for increasing I until the point is reached where $I = J_S(0)$, or $\Phi = \pi/2$. Then the current jumps from the zero-voltage line to the right and on to the experimental curve for single-particle tunneling, which is shown as the dashed line. In fact, in most tunnel junctions, the maximum phase $\Phi = \pi/2$ cannot be reached, and the jump to the single-particle tunneling curve occurs at applied currents smaller than the maximum $J_S(0)$.

Another experiment is to take a tunnel junction between two superconductors and to apply a static voltage $eV < 2\Delta$. Then one observes an oscillatory current according to (10.124). This oscillatory current is the tunneling of pairs from the chemical potential of one superconductor to the chemical potential of the other. This tunneling of pairs does not conserve energy but takes an energy $2eV$ for the pair of charge $2e$ to overcome the voltage V. The pair must tunnel back, since it cannot complete a transition where energy is not conserved. This ac current can be observed by the external radiation it produces (Langenburg *et al.*, 1965).

The preceding two experiments must be performed in zero static magnetic field. The Josephson effect is so sensitive to magnetic field that care must be used to cancel out the effects of the earth's magnetic field. This sensitivity to magnetic field, which was pointed out by Josephson (1962), is useful for verifying that the tunneling is due to the Josephson effect

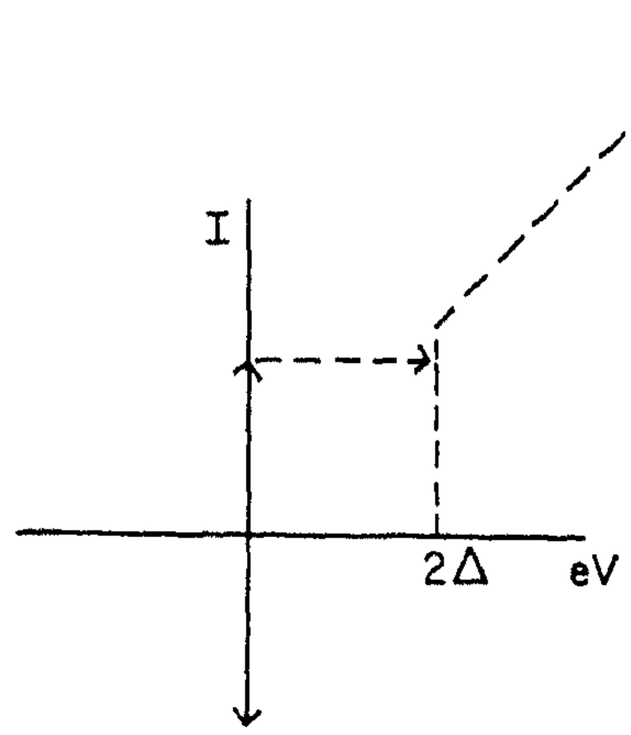


FIGURE 10.13 Current–voltage characteristics of a Josephson junction. The dashed line rising abruptly at $eV = 2\Delta$ is the single-particle tunneling curve in Fig. 10.10 at $T = 0$. The arrows on the vertical axis show that a current can flow with no voltage. But an imposition of too much current causes the device to jump to the single-particle curve.

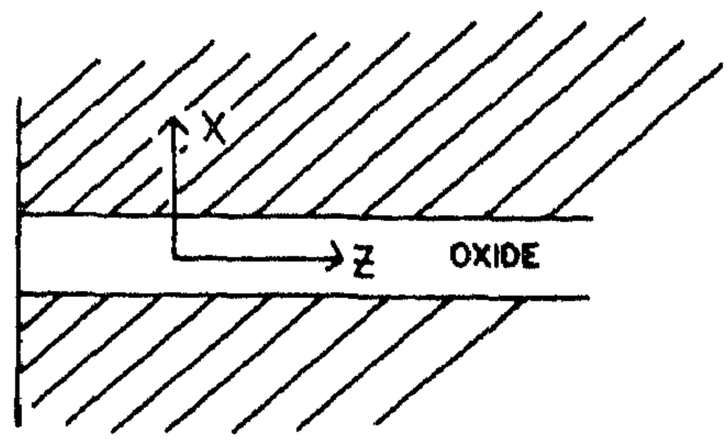


FIGURE 10.14 Coordinate system used in discussing Josephson tunneling.

and not some other process such as a superconducting short in the oxide—the latter would be expected to behave like the current characteristic in Fig. 10.13. The method of including magnetic field is as follows. The total phase difference between the superconductors is written as Φ. The phase is now meant to include the applied voltage V, so that we write it as

$$I_J = J_S(eV)\sin(\Phi) \tag{10.132}$$

$$\Phi = \Phi_0 + \frac{2e}{\hbar}\int_0^t dt' \int_a^b dx \frac{dV}{dx} \tag{10.133}$$

The spatial coordinates are shown in Fig. 10.14. The integral $\int dx$ in (10.133) is taken across the junction from one side to the other. The dx integral is actually $V(b) - V(a) \equiv V$, since a and b are on two different sides of the oxide and each is deep within the bulk of the superconductor.

The phase Φ needs to be expressed in a manner which is gauge invariant. Write the phase as an electric field $\mathbf{E}(\mathbf{r}, t)$ which is always gauge invariant (see Sec. 1.5):

$$\frac{dV}{dx} \Rightarrow -E_x = \frac{1}{c}\frac{\partial A_x}{\partial t} + \frac{\partial \phi}{\partial x} \tag{10.134}$$

$$\Phi = \Phi_0 + \frac{2e}{\hbar c}\int dx' \int dt' \frac{\partial A_x}{\partial t'} + \frac{2e}{\hbar}\int dx' dt' \frac{\partial \phi}{\partial x'} \tag{10.135}$$

The vector potential A_x term is manipulated in the following fashion. It is split into time-dependent and time-independent terms. The time-independent term is obtained by doing the time integral

$$\int^t dt' \frac{\partial A_x}{\partial t'} = A_x(t) = A_x \tag{10.136}$$

Second, the space integral $\int dx' A_x$ is recognized as a line integral $\int d\mathbf{l} \cdot \mathbf{A}$ along a particular path, and any path will give the same result. Then the use of Stokes' theorem brings us to

$$\int d\mathbf{l} \cdot \mathbf{A} = \int d\mathbf{s} \cdot \nabla \times \mathbf{A} = \int d\mathbf{s} \cdot \mathbf{B}_0 \tag{10.137}$$

$$\Phi = \Phi_0 + \frac{2e}{\hbar c}\int d\mathbf{s} \cdot \mathbf{B}_0 - \frac{2e}{\hbar}\int^t dt' \int d\mathbf{l}' \cdot \mathbf{E}(\mathbf{r}', t') \tag{10.138}$$

The identity $\nabla \times \mathbf{A} = \mathbf{B}_0$ gives the term $\int d\mathbf{s} \cdot \mathbf{B}_0$, which is the magnetic flux under the area of the integral. The external magnetic field $\mathbf{B}_0 = B_0\hat{y}$ is constant when $\mathbf{A}$ is constant. The area $d\mathbf{s}$ is the distance z from some reference point, which can be the end of the junction, times the effective distance in the x direction. The latter is the thickness of the oxide d plus the distance

the magnetic field penetrates into the superconductor on each side of the junction. The latter distances are the two penetration depths $\lambda_R + \lambda_L$. In a fixed magnetic field the phase is

$$\Phi = \Phi_0 + kz + \omega t \tag{10.139}$$

$$k = \frac{2e}{\hbar c} B_0 (d + \lambda_R + \lambda_L) \tag{10.140}$$

The magnetic field generates an effective wave vector k. The Josephson current I_J now depends on the distance z from the reference point. The total current in the sample of length L in the z direction is found by integrating in this direction ($eV = 0$):

$$j(z) = J_S(0) \sin(\Phi_0 + kz) \tag{10.141}$$

$$I_J = \frac{1}{L} \int_{-L/2}^{L/2} dz j(z) = J_S(0) \sin(\Phi_0) \left[\frac{2}{kL} \sin\left(\frac{kL}{2}\right) \right] \tag{10.142}$$

The theoretical prediction is that the maximum current at zero voltage has a Fraunhofer pattern $\sin(\theta)/\theta$, where $\theta \propto B_0$. The Fraunhofer pattern was indeed observed by Fiske, as shown in Fig. 10.15 for Sn–oxide–Sn junctions. This experiment is another spectacular verification of Josephson's predictions. Even more interesting results are obtained when there is both a static dc magnetic field B_0 and dc voltage V. Then the Josephson current has the form

$$j(z) = J_S(eV) \sin(\Phi_0 + kz + \omega t) \tag{10.143}$$

This form of current is also observed in a variety of experiments which are described in the review volume edited by Parks (1969).

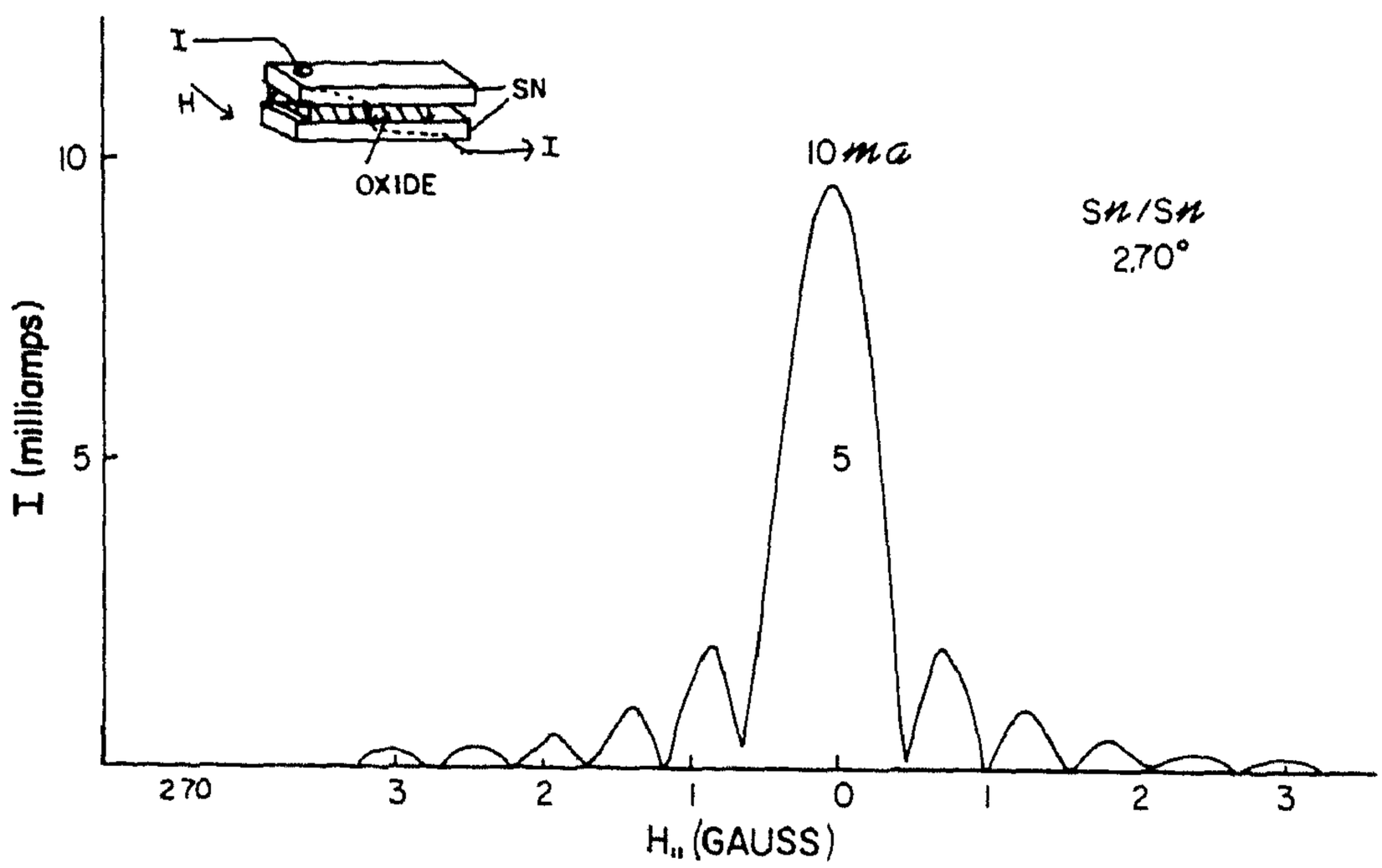


FIGURE 10.15 The Fraunhofer pattern of a Josephson junction. The experimental curve is the maximum dc current which can flow with no voltage as a function of applied magnetic field. *Source:* Fiske (1978, unpublished) (used with permission).

Another way to evaluate the integral in (10.142) is to separate the negative values of z, and then to change $z' = -z$ in this integrand

$$I_J = \frac{J_S(0)}{L}\left[\int_{-L/2}^{0} dz \sin(\Phi_0 + kz) + \int_0^{L/2} dz \sin(\Phi_0 + kz)\right] \tag{10.144}$$

$$= \frac{J_S(0)}{L}\left[\int_0^{L/2} dz \sin(\Phi_0 - kz) + \int_0^{L/2} dz \sin(\Phi_0 + kz)\right] \tag{10.145}$$

where the first integral is for the left side of the junction, and the second integral is for the right side.

A dc Josephson current at zero magnetic field is observed between lead and the high temperature cuprates. This observation is evidence that the cuprates are also spin singlet states. A Josephson current is not allowed between two different superconductors if one has the spins in a singlet and the other has the spins in a triplet. This experiment proves that the cuprates are not p-wave superconductors, since symmetry demands that the p-state must have a spin triplet. See Chapter 11 for a discussion of p-wave superconductivity.

The same Fraunhofer pattern is observed for flat junctions, if one side has lead and the other side has a cuprate superconductor. Wollman *et al.* (1995) made a tunnel junction that wrapped around a ninety degree corner. The junction now has the geometry shown in Fig. 10.16(a). The inside corner is a cuprate superconductor, while the outside is lead. It is known that lead is an s-wave superconductor. If the cuprates are also s-wave, one should get the same Fraunhofer pattern as for a flat junction. However, if the cuprates have d-waves symmetry, the tunneling currents in the two arms of the corner should be out of phase. This case can be

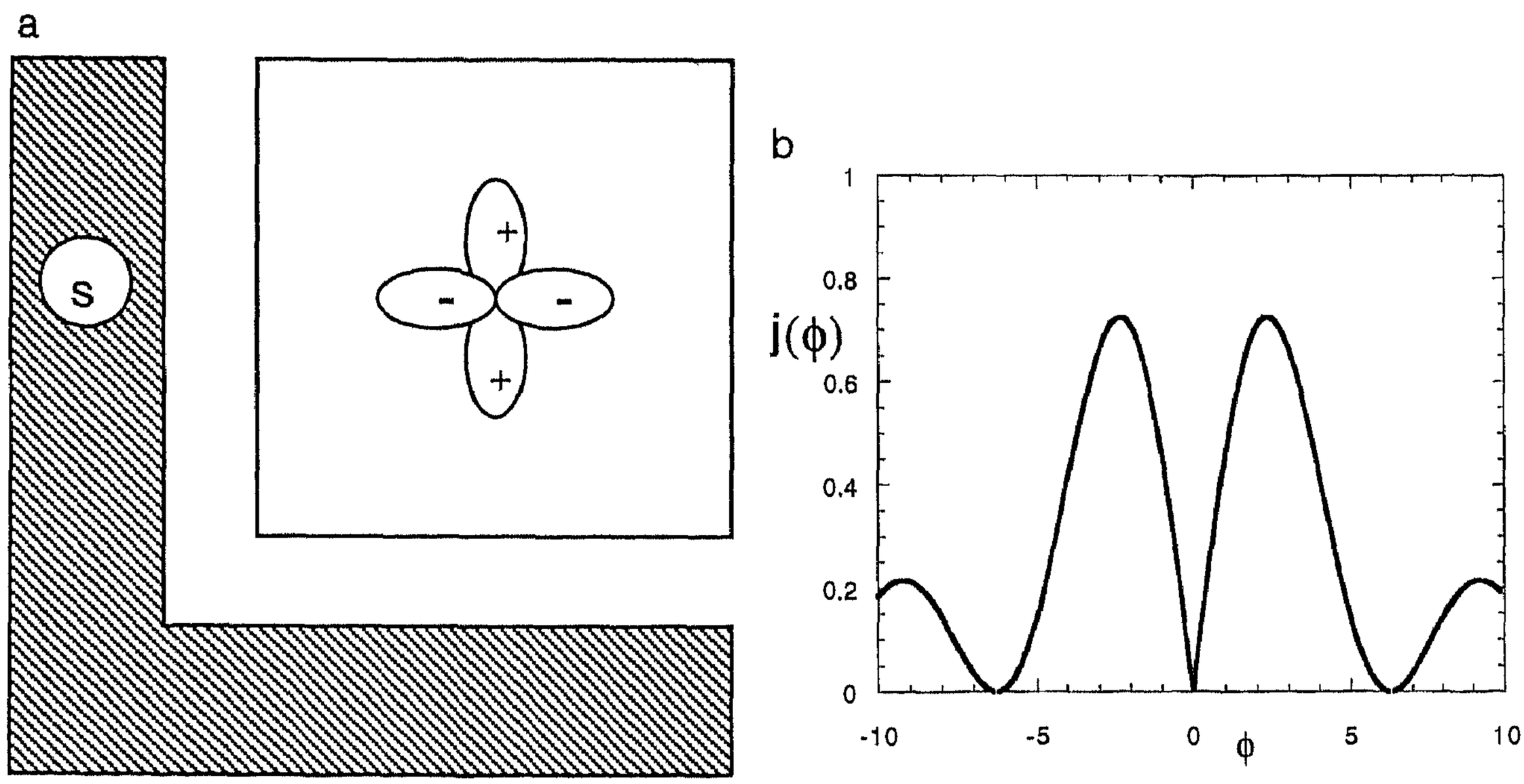


FIGURE 10.16 (a) Geometry of a corner junction between s and d-wave superconductors. (b) Critical current vs. applied magnetic field for a corner junction. The predicted current is zero at zero magnetic field since the currents are in opposite directions in the two arms.

evaluated using (10.145), where the two integrals represent the two arms of the corner. Changing the relative sign between the two integrals gives

$$(I_J)_{sd} = J_S(0)\cos(\Phi_0)\left(\frac{1-\cos\theta}{\theta}\right) \tag{10.146}$$

$$\theta = \frac{kL}{2} \tag{10.147}$$

The pattern of maximum current is no longer Fraunhofer. The predicted behavior of a corner junction is shown in Fig. 10.16(b). Precisely this behavior was found for corner junctions by Wollman *et al.* (1995). The Josephson current vanishes at zero magnetic field because the current is in a different direction in each arm. The two arms have to have the same length for perfect cancellation. Their experimental result is strong evidence for d-wave superconductivity in the cuprates.

10.2.4. Infrared Absorption

A unique feature of the BCS theory was the prediction of an energy gap in the excitation spectrum of an s-wave superconductor at zero temperature. Electron tunneling (Giaever, 1960) was the first measurement to convincingly demonstrate the existence of the energy gap. However, this experiment was not commonplace at the time of the BCS theory (1957), and the conventional method of seeking an energy gap would be to measure the optical absorption of the metal in the far infrared. For example, Pb has $T_c = 7.2\,\mathrm{K}$ so $k_B T_c = 0.62\,\mathrm{meV}$. The BCS prediction is that the energy gap is $E_g = 2\Delta = 3.52(k_B T_c) = 2.2\,\mathrm{meV} = 17.7\,\mathrm{cm}^{-1}$. The value found experimentally is $E_g = 22.5\,\mathrm{cm}^{-1}$. The difference from the BCS prediction is due to the fact that Pb is a strongly coupled superconductor. The frequency range is very difficult experimentally so that good experimental results for Pb were not reported until 1968 (Palmer and Tinkham, 1968).

An s-wave superconductor at zero temperature should not absorb radiation for $\omega < 2\Delta$. The absorption breaks up an electron pair, which creates two quasiparticles. Each quasiparticle has an energy of the form $E = \sqrt{\xi^2 + \Delta^2}$, so absorption starts at $\hbar\omega = 2\Delta$. The theoretical discussion follows the original theory by Mattis and Bardeen (1958).

The rate of absorption in the infrared is given by the real part of the conductivity, which is evaluated from the Kubo formula:

$$\pi(\mathbf{q}, i\omega) = -\frac{1}{3\mathrm{v}}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \mathbf{j}(\mathbf{q},\tau)\cdot\mathbf{j}(-\mathbf{q},0)\rangle \tag{10.148}$$

$$\sigma(\mathbf{q}, i\omega) = \frac{1}{\omega}\mathrm{Im}[\pi_{\mathrm{ret}}(\mathbf{q},\omega)] \tag{10.149}$$

$$\mathbf{j}(\mathbf{q}) = \frac{e}{m}\sum_{\mathbf{p}\sigma}(\mathbf{p}+\tfrac{1}{2}\mathbf{q})C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}C_{\mathbf{p}\sigma} \tag{10.150}$$

The expression for the current is appropriate for noninteracting particles. There results a correlation function of four operators

$$\pi(\mathbf{q}, i\omega) = -\frac{e^2}{3m^2 v} \sum_{\mathbf{p}\mathbf{p}'\sigma\sigma'} (\mathbf{p} + \tfrac{1}{2}\mathbf{q}) \cdot (\mathbf{p}' - \tfrac{1}{2}\mathbf{q})$$
$$\times \int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau C^\dagger_{\mathbf{p}+\mathbf{q},\sigma}(\tau) C_{\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma}(0) C_{\mathbf{p}'\sigma'}(0)\rangle \tag{10.151}$$

In prior discussions of optical absorption processes, it was convenient to set to zero the photon wave vectors $\mathbf{q}$. The step $\mathbf{q} = 0$ is not done here because later it would be found that the conductivity is zero. It is formally necessary to retain a nonzero value for $\mathbf{q}$, although the magnitude of $q \approx 10^2\,\mathrm{cm}^{-1}$ can certainly be neglected compared to the wave vectors of the electrons $k_F \approx 10^8\,\mathrm{cm}^{-1}$.

To evaluate the bracket of operators in (10.151), fix one spin, say $\sigma = \uparrow$. The answer is twice this result, since the same contribution is found for $\sigma = \downarrow$. The operators in the angle brackets can be paired two ways (the pairings with $\mathbf{q} = 0$ are not allowed since $\mathbf{q}$ is nonzero):

$$\langle T_\tau C^\dagger_{\mathbf{p}+\mathbf{q},\uparrow}(\tau) C_{\mathbf{p}\uparrow}(\tau) C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma'}(0) C_{\mathbf{p}'\sigma'}(0)\rangle = -\delta_{\sigma'=\uparrow}\delta_{\mathbf{p}'=\mathbf{p}+\mathbf{q}} \mathscr{G}(\mathbf{p}, \tau)\mathscr{G}(\mathbf{p}+\mathbf{q}-\tau) \tag{10.152}$$
$$+ \delta_{\sigma'=\downarrow}\delta_{\mathbf{p}'=-\mathbf{p}} \mathscr{F}(\mathbf{p}, \tau)\mathscr{F}^\dagger(\mathbf{p}+\mathbf{q}, \tau) \tag{10.153}$$

These pairings evaluate the correlation function, which is written in both τ-space and Fourier space:

$$\pi(\mathbf{q}, i\omega) = \frac{2e^2}{3m^2 v} \sum_{\mathbf{p}} (\mathbf{p} + \tfrac{1}{2}\mathbf{q})^2 \int_0^\beta d\tau e^{i\omega_n \tau}$$
$$\times [\mathscr{G}(\mathbf{p}, \tau)\mathscr{G}(\mathbf{p}+\mathbf{q}, -\tau) + \mathscr{F}(\mathbf{p}, -\tau)\mathscr{F}^\dagger(\mathbf{p}+\mathbf{q}, \tau)] \tag{10.154}$$

$$\pi(\mathbf{q}, i\omega) = \frac{2e^2}{3m^2 v} \sum_{\mathbf{p}} (\mathbf{p} + \tfrac{1}{2}\mathbf{q})^2 \frac{1}{\beta} \sum_{ip}$$
$$\times [\mathscr{G}(\mathbf{p}, ip)\mathscr{G}(\mathbf{p}+\mathbf{q}, ip+\omega) + \mathscr{F}(\mathbf{p}, ip)\mathscr{F}^\dagger(\mathbf{p}+\mathbf{q}, ip+i\omega)] \tag{10.155}$$

Note that the vector vertex of the current operator changes the relative minus sign between the two terms in (10.152) to a relative plus in (10.155). The two terms in (10.155) can be represented by the two diagrams shown in Fig. 10.17. The photon of $(\mathbf{q}, \omega)$ enters and leaves, and couples to the two solid lines of the polarization diagram. The solid lines can be either the pair $\mathscr{G}\mathscr{G}$ or $\mathscr{F}\mathscr{F}^\dagger$ as these are the only two possibilities which conserve particle number at the two vertices. One could not, for example, have the combination $\mathscr{G}\mathscr{F}$ since that does not conserve particle number at one of the vertices. The crucial aspect of the above expression is the relative sign between the two terms $\mathscr{G}\mathscr{G}$ and $\mathscr{F}\mathscr{F}^\dagger$. There seems to be no easy way to deduce this sign, except by being careful.

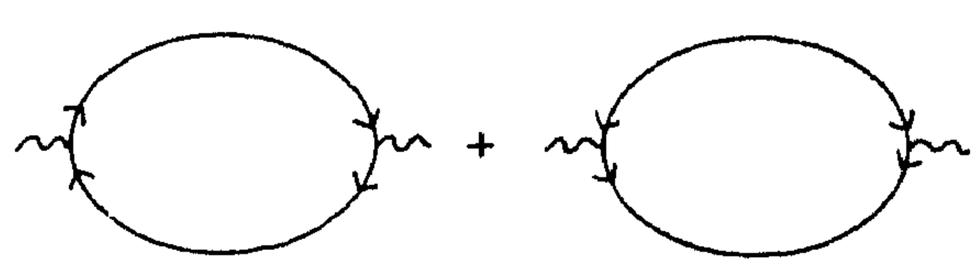


FIGURE 10.17

The next step in the derivation is the summation over Matsubara frequencies. The summation for $\mathscr{F}\mathscr{F}^\dagger$ was given previously in (10.117), while the other one is

$$\begin{aligned}\frac{1}{\beta}\sum_{ip}\mathscr{G}(\mathbf{p}, ip)\mathscr{G}(\mathbf{k}, ip+i\omega)\\ = [1-n_F(E_p)-n_F(E_k)]\left(\frac{u_k^2v_p^2}{i\omega-E_p-E_k}-\frac{u_p^2v_k^2}{i\omega+E_p+E_k}\right)\\ +[n_F(E_p)-n_F(E_k)]\left(\frac{u_k^2u_p^2}{i\omega+E_p-E_k}-\frac{v_p^2v_k^2}{i\omega-E_p+E_k}\right)\end{aligned} \tag{10.156}$$

At nonzero temperatures, there are a number $n_F(E)$ of quasiparticle excitations with energy E which can absorb the radiation even for $\hbar\omega < 2\Delta$. The interesting case is the limit of zero temperature, where there is no absorption until $\hbar\omega > 2\Delta$. At zero temperature the nonzero terms are

$$\begin{aligned}\pi(\mathbf{q}, i\omega) = \frac{2e^2}{3m^2\mathrm{v}}\sum_{\mathbf{p}}(\mathbf{p}+\tfrac{1}{2}\mathbf{q})^2\\ \times\left(\frac{u_{\mathbf{p}+\mathbf{q}}^2v_{\mathbf{p}}^2-u_{\mathbf{p}}v_{\mathbf{p}}u_{\mathbf{p}+\mathbf{q}}v_{\mathbf{p}+\mathbf{q}}}{i\omega-E_{\mathbf{p}}-E_{\mathbf{p}+\mathbf{q}}}-\frac{u_{\mathbf{p}}^2v_{\mathbf{p}+\mathbf{q}}^2-u_{\mathbf{p}}v_{\mathbf{p}}u_{\mathbf{p}+\mathbf{q}}v_{\mathbf{p}+\mathbf{q}}}{i\omega+E_{\mathbf{p}}+E_{\mathbf{p}+\mathbf{q}}}\right)\end{aligned} \tag{10.157}$$

The retarded function is obtained from $i\omega \to \omega + i\delta$, and taking the imaginary part of the retarded function brings us to the real part of the conductivity as $\mathrm{v} \to \infty$ for $\omega > 0$:

$$\sigma(\mathbf{q}, \omega) = \frac{2\pi e^2}{3m^2\omega}\int\frac{d^3p}{(2\pi)^3}(\mathbf{p}+\tfrac{1}{2}\mathbf{q})^2[u_{\mathbf{p}+\mathbf{q}}^2v_{\mathbf{p}}^2-u_{\mathbf{p}}v_{\mathbf{p}}u_{\mathbf{p}+\mathbf{q}}v_{\mathbf{p}+\mathbf{q}}]\delta(\omega-E_{\mathbf{p}}-E_{\mathbf{p}+\mathbf{q}}) \tag{10.158}$$

This expression has a reasonable form. The photon is absorbed by having its energy $\hbar\omega$ converted into two quasiparticles with separate energies $E_{\mathbf{p}}$ and $E_{\mathbf{p}+\mathbf{q}}$.

The combination of coherence factors $u_{\mathbf{p}+\mathbf{q}}^2v_{\mathbf{p}}^2 - u_{\mathbf{p}}v_{\mathbf{p}}u_{\mathbf{p}+\mathbf{q}}v_{\mathbf{p}+\mathbf{q}}$determines the rate of absorption. This quantity vanishes as $\mathbf{q} \to 0$. It was remarked that $q \ll p$, so that one should indeed take this limit. Then one would conclude that the absorbing power of the superconductor was very weak, at least for the process of creating two quasiparticles. This conclusion is valid in the case where momentum is a good quantum number, so the two quasiparticles made by the photon have momenta which are related, i.e., $\mathbf{p}$ and $\mathbf{p}+\mathbf{q}$.

Mattis and Bardeen assumed that momentum was not conserved. Instead, the two excited quasiparticles have momentum $\mathbf{p}$ and $\mathbf{p}'$ which are unrelated. The lack of momentum conservation is characteristic of the so-called *dirty superconductor*. The momentum breaking might occur, for example, because the experiments are done on evaporated metal films which are not crystalline. Alternately, there might be phonons or impurities which scatter the particles. In any case, it is assumed that energy is conserved but not momentum. Separately sum over $\mathbf{p}$ and $\mathbf{p}'$ which leads to the expression

$$\sigma(\omega) = \frac{C_0}{\omega}\int_{-\infty}^{\infty}d\xi_p\int_{-\infty}^{\infty}d\xi_p'\delta(\omega-E_p-E_{p'})[v_p^2u_{p'}^2-u_pv_pu_{p'}v_{p'}]$$

where C_0 is a constant which will be determined later. The terms in the coherence factors v_p^2 and u_p^2, which are linear in ξ_p and $\xi_{p'}$, vanish because they are odd functions of ξ_p and $\xi_{p'}$.

Then change variables to $E = E_p$ and $E' = E_{p'}$, and derive the expression

$$\sigma(\omega) = \frac{C_0}{\omega}\int_\Delta^\infty dE\rho(E)\int_\Delta^\infty dE'\rho(E')\delta(\omega - E - E')\left(1 - \frac{\Delta^2}{EE'}\right) \tag{10.159}$$

The E' integral is done by energy conservation:

$$\sigma(\omega) = \frac{C_0}{\omega}\Theta(\omega - 2\Delta)\int_\Delta^{\omega-\Delta} dE\rho(E)\rho(\omega - E)\left(1 - \frac{\Delta^2}{E(\omega - E)}\right)$$

This integral has the same form as (10.105) and is evaluated in the same manner: change variables of integration to x, where $2E = \omega + x(\omega - 2\Delta)$, and the integral has the form for complete elliptic integrals:

$$\alpha = \frac{\omega - 2\Delta}{\omega + 2\Delta} \tag{10.160}$$

$$\begin{aligned}\sigma(\omega) &= \frac{C_0}{2\omega}\Theta(\omega - 2\Delta)(\omega - 2\Delta)\int_{-1}^{1} dx\frac{1 - \alpha x^2}{[(1 - x^2)(1 - \alpha^2 x^2)]^{1/2}} \\ \sigma(\omega) &= \frac{C_0}{\omega}\Theta(\omega - 2\Delta)[\omega + 2\Delta)E(\alpha) - 4\Delta K(\alpha)]\end{aligned} \tag{10.161}$$

The absorption for a normal metal has $\Delta = 0$, which in the same units is C_0. The ratio of the conductivity in the superconductor to that of the normal metals is

$$\frac{\sigma_s(\omega)}{\sigma_n(\omega)} = \frac{1}{\omega}\Theta(\omega - 2\Delta)[(\omega + 2\Delta)E(\alpha) - 4\Delta K(\alpha)] \tag{10.162}$$

which is the result of Mattis and Bardeen.

The experiment measures the conductivity in the normal metal and the superconductor at the same temperature. The metal is made normal, at the low temperature, by the application of a small magnetic field. The experimental results of Palmer and Tinkham (1968) for Pb are shown as the points in Fig. 10.18. They compare well with the theory of Mattis and Bardeen, Eq. (10.162), which is shown as the solid line. There is no absorption until the energy gap of 22.5 cm^{-1}. Then the absorption rises gradually with increasing ω. It does not, as in the case of tunneling between two superconductors, rise discontinuously at the threshold frequency.

The agreement between theory and experiment is obviously excellent. In fact it is initially surprising that the agreement is that good. After all, Pb is a strongly coupled superconductor, and there is no particular reason to expect the BCS theory to work well for this metal. This question was investigated by Shaw and Swihart (1968), who solved the theory of infrared absorption using the strong coupling theory. They found that the results were nearly identical to the predictions of the BCS theory. This experiment is not sensitive to the strength of the coupling. There are, however, numerous other experiments which can distinguish between these theories.

Many far-infrared experiments have been performed on high-temperature superconductors. The results have generally been inconclusive. There is no peak which can be identified as due to the energy gap of the superconductors. The cuprates are oxides, and have many optical phonons in the same range of frequency, which complicates the interpretation. Of course, if the order parameter has d-wave character, then there is no energy gap in the infrared. The node in the gap function means that there is a distribution of energy gaps, which extends down to zero energy.

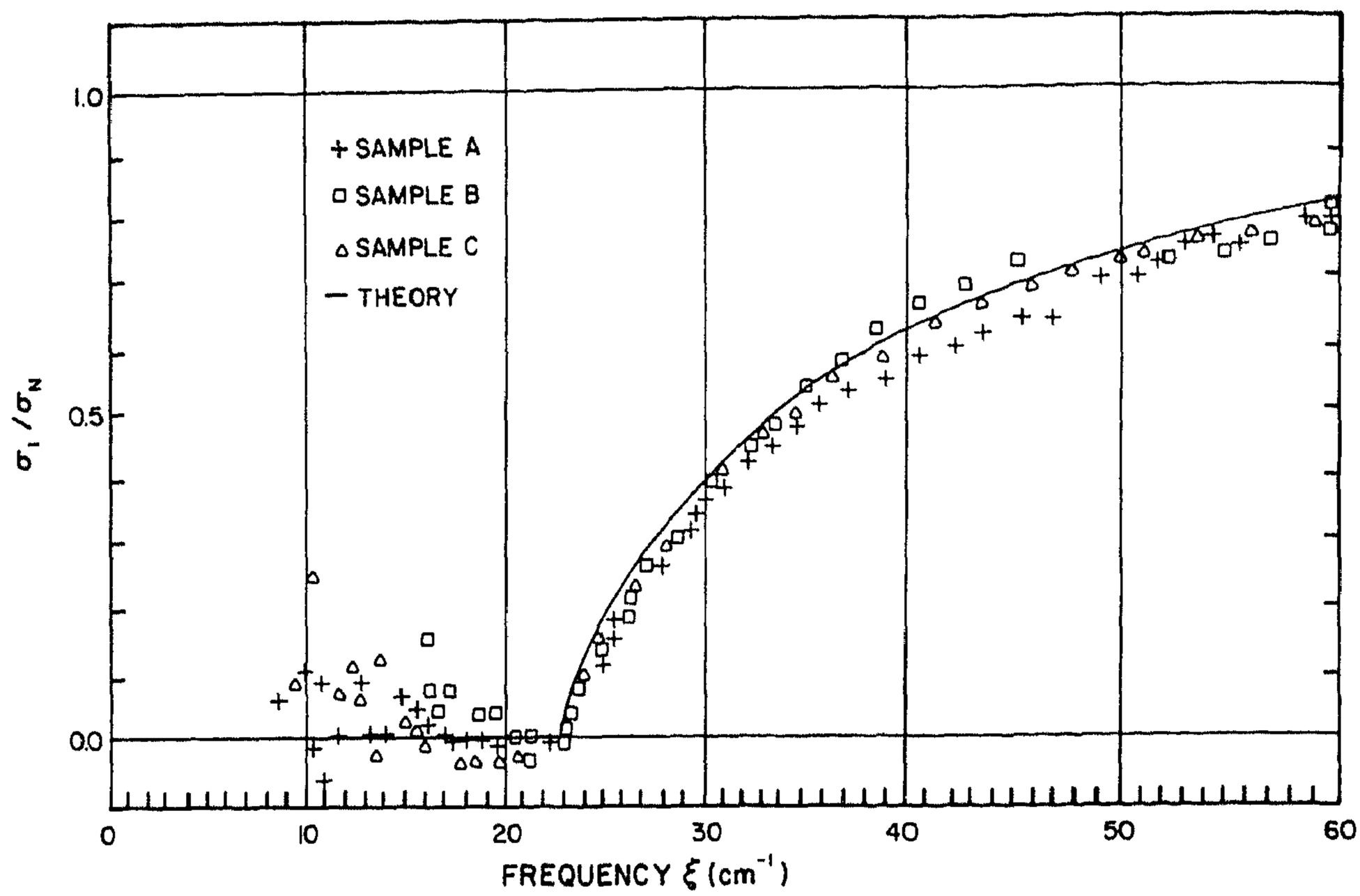


FIGURE 10.18 Far-infrared absorption in superconducting lead at 2 K. The solid line is BCS theory, and points are experimental. Source: Palmer and Tinkham (1968) (used with permission).

10.3. STRONG COUPLING THEORY

The BCS theory is based on a simple approximation (10.2) to the attractive interaction between two electrons. This theory successfully describes the properties of many elemental superconductors. Nevertheless, there are some superconductors for which the electron–phonon coupling is quite strong, and here the BCS theory is not an accurate description of the superconducting state. Instead, one must solve the gap equation for a realistic interaction between two electrons, which includes both the attraction due to phonons and the repulsion due to the screened Coulomb interaction. It is important that the phonon term be fully retarded, i.e., that the interaction be frequency dependent. This strong coupling theory was described by Eliashberg (1960). One of its important features is that the gap function $\Delta(\omega)$ becomes a complex function of the real frequency ω. Alternately, it is a real function $\Delta(ip_n)$ of the imaginary frequency p_n. In addition, the self-energy function of the electrons, due to electron–phonon interaction, must be retained and plays an important role in the analysis. The discussion of strong coupling theory follows Scalapino *et al.* (1966).

The starting point in the theory is the screened interaction between two electrons, which is written in the conventional way as the sum of a screened Coulomb interaction and a screened phonon interaction (see Secs. 7.3 and 7.4):

$$V_{\text{eff}}(\mathbf{q}, i\omega) = V_e(\mathbf{q}) + \sum_\lambda \bar{M}_\lambda^2(\mathbf{q})\mathscr{D}_\lambda(\mathbf{q}, i\omega) \tag{10.163}$$

$$V_e(\mathbf{q}) = \frac{4\pi e^2}{q^2\varepsilon(\mathbf{q})} \tag{10.164}$$

$$\bar{M}_\lambda(\mathbf{q}) = \frac{M_\lambda(\mathbf{q})}{\varepsilon(q)} \tag{10.165}$$

The Coulomb term V_e is taken to be instantaneous, since the frequency dependence of the dielectric function $\varepsilon(\mathbf{q}, \omega)$ is not important for the low frequencies of interest, $\omega \leq \omega_D$. Similarly, the screened electron–phonon coupling constant $\bar{M}_\lambda(\mathbf{q})$ is also taken to be a function of wave vector but not frequency. These were the same approximations used in Sec. 7.4 to discuss the electron–phonon effects in normal metals.

The objective is to calculate, self-consistently, the properties of the correlation functions $\mathscr{G}$, $\mathscr{F}$, and $\mathscr{F}^\dagger$. The last two are again equal, so it is sufficient to find one of them. They have the same definition (10.31) as before. Now they will be evaluated by directly expanding the S matrix and summing the terms which are found in higher order:

$$\mathscr{G}(\mathbf{p}, \tau - \tau') = -\langle T_\tau C_{\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle = -\langle T_\tau \hat{C}_{\mathbf{p}\sigma}(\tau) S \hat{C}^\dagger_{\mathbf{p}\sigma}(\tau')\rangle$$

$$\mathscr{F}(\mathbf{p}, \tau - \tau') = \langle T_\tau C_{-\mathbf{p}\downarrow}(\tau) C_{\mathbf{p}\uparrow}(\tau')\rangle = \langle T_\tau \hat{C}_{-\mathbf{p}\downarrow}(\tau) S \hat{C}_{\mathbf{p}\uparrow}(\tau')\rangle$$

$$\mathscr{F}^\dagger(\mathbf{p}, \tau - \tau') = \langle T_\tau C^\dagger_{\mathbf{p}\uparrow}(\tau) C^\dagger_{-\mathbf{p}\downarrow}(\tau')\rangle = \langle T_\tau \hat{C}^\dagger_{\mathbf{p}\uparrow}(\tau) S \hat{C}^\dagger_{-\mathbf{p}\downarrow}(\tau')\rangle$$

Two different self-energy functions are retained in this expansion. Both are one-phonon self-energies, where the electron line in the self-energy can be either $\mathscr{G}$ or $\mathscr{F}$:

$$S(\mathbf{p}, ip) = -\int \frac{d^3q}{(2\pi)^3} \frac{1}{\beta} \sum_{iq} V_{\text{eff}}(\mathbf{q}, iq) \mathscr{G}(\mathbf{p} + \mathbf{q}, ip + iq) \tag{10.166}$$

$$W(\mathbf{p}, ip) = -\int \frac{d^3q}{(2\pi)^3} \frac{1}{\beta} \sum_{iq} V_{\text{eff}}(\mathbf{q}, iq) \mathscr{F}(\mathbf{p} + \mathbf{q}, ip + iq) \tag{10.167}$$

These self-energies are shown as Feynman diagrams in Fig. 10.19. The dashed line is V_{eff} in (10.163). The electron line, in the self-energy expression, contains the Green's functions $\mathscr{G}$ or $\mathscr{F}$ which are to be determined self-consistently. This theory neglects the vertex corrections, which are the self-energy diagrams of the type shown in Fig. 10.19(c). This neglect is customarily justified on the basis of a theorem due to Migdal (1958), who argued that such terms were smaller by a factor of $\sqrt{m/M}$, where m is the electron mass and M is the ion mass. The difficulty with his argument is that superconductivity itself (i.e., the Cooper instability) is caused by a vertex diagram. A theorem which asserts that all vertex diagrams may be neglected is obviously unreliable. The strong coupling theory is described in its customary form, which does neglect these vertex corrections.

First discuss the wave vector dependence of $W(\mathbf{p}, ip_n)$. The wave vectors of interest are on the Fermi surface. Since the Fermi surface is not a sphere, the magnitude of the Fermi wave vector will depend upon the direction in the crystal. For an isotropic s-wave superconductor, then just ignore the variation in the Fermi wave vector. This approximation is tantamount to neglecting the wave vector dependence, and writing the gap as $W(ip_n)$. There are many cases where this approximation is poor, and one should include the dependence of the gap upon direction in the crystal. Because of crystal symmetry, this dependence can usually be accomplished with a few functions, which are the basis functions of the group

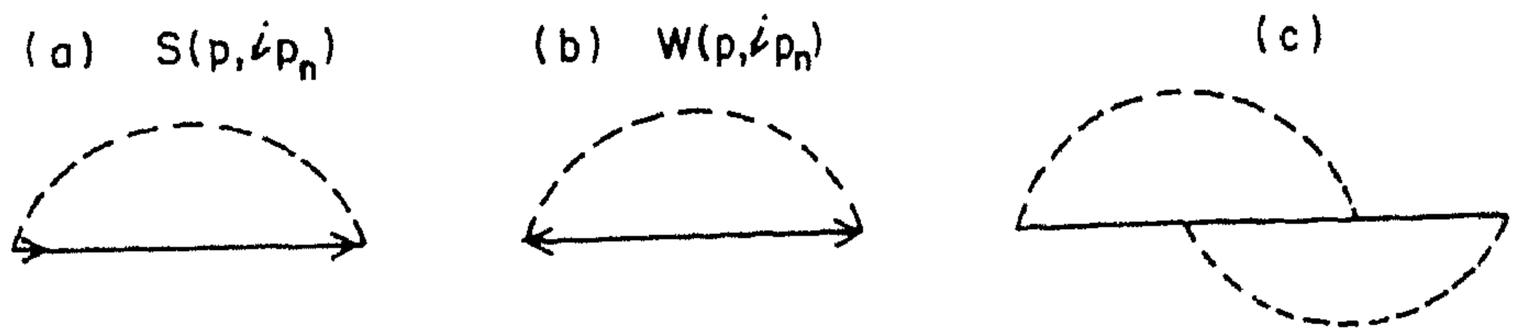


FIGURE 10.19

which describes the crystal symmetry. For cubic crystals these are called Kubic Harmonics (van der Lage and Bethe, 1947). For other crystals, similar functions can be constructed. For example, in high-temperature superconductors, recent experiments have indicated that the gap may have d-wave symmetry. The symmetry is $x^2 - y^2$. Define the angle ϕ for the two-dimensional square lattice by $\tan(\phi) = k_y/k_x$. Then a good approximation to the gap function is $W(\mathbf{p}, ip_n) = W_2(ip_n)\cos(2\phi)$. The angular function $\cos(2\phi)$ gives the proper symmetry of being plus one along the $\pm x$ axes and minus one along the $\pm y$ axes. This ansatz form could be used in the gap equation. In this case it is relatively easy to integrate out the angle ϕ so that one is left again with solving a matrix equation where the only important variable is the imaginary frequency.

Normally the expansion of the S matrix leads to a Dyson equation of the typical form:

$$\mathscr{G}(p) = \mathscr{G}^{(0)}(p)[1 + \Sigma(p)\mathscr{G}(p)] \tag{10.168}$$

The equivalent expansion is derived for the superconductor. The situation is obviously more complicated, since there are two self-energy functions $S(p)$ and $W(p)$, and two Green's functions $\mathscr{G}$ and $\mathscr{F}$. The noninteracting function $\mathscr{G}^{(0)}(p) = 1/(ip - \xi_p)$ as usual, but obviously there is not an equivalent form for $\mathscr{F}^{(0)}$, so $\mathscr{F}^{(0)} = 0$. The relevant Dyson's equations are, in a four-vector notation $p \equiv (\mathbf{p}, ip)$

$$\begin{aligned} \mathscr{G}(p) &= \mathscr{G}^{(0)}(p)[1 + S(p)\mathscr{G}(p) - W(p)\mathscr{F}^{\dagger}(p)] \\ \mathscr{F}(p) &= \mathscr{G}^{(0}(-p)[W(p)\mathscr{G}(p) + S(-p)\mathscr{F}^{\dagger}(p)] \end{aligned} \tag{10.169}$$

These equations are shown diagrammatically in Fig. 10.20. The key to constructing the equations is to remember that the vertex conserves particle number, so that the two electron arrows must point in the same direction; i.e., one can only have vertices of the form below. These equations are simplified because $W(-p) = W(p)$. Dyson's equations are quite similar

to the BCS equations (10.56). This resemblance is made more apparent by rewriting (10.169) using $\mathscr{G}^{(0}(\pm p) = 1/(\pm ip - \xi_p)$ to obtain

$$[ip - \xi_p - S(p)]\mathscr{G} + W(p)\mathscr{F}^{\dagger} = 1 \tag{10.170}$$

$$[ip + \xi_p + S(-p)]\mathscr{F}^{\dagger} + W(p)\mathscr{G} = 0 \tag{10.171}$$

The BCS equations (10.56) are obtained in the approximation where $W(\mathbf{p}, ip) = \Delta =$ constant, and the other self-energy $S(\mathbf{p},ip)$ is ignored. The more general equations (10.171) are now solved. The gap function $W(p)$ is an even function of ip, as already stated

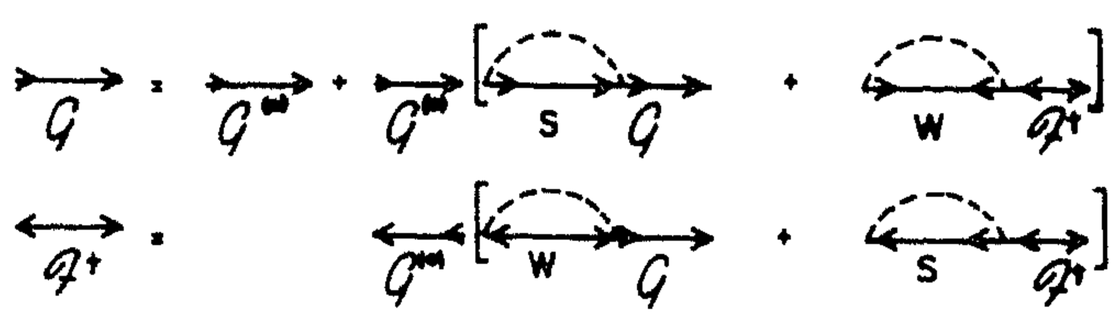


FIGURE 10.20

above. The other self-energy $S(p)$ is not. It is convenient to break it up into its symmetric and antisymmetric parts:

$$S(p) = S_e(\mathbf{p}, ip) + ipS_o(\mathbf{p}, ip) \tag{10.172}$$

where S_e and S_o are both even functions of frequency ip. Furthermore, the function $Z(\mathbf{p}, ip)$ is defined as

$$Z(\mathbf{p}, ip) = 1 - S_o(\mathbf{p}, ip) \tag{10.173}$$

The notation Z usually is applied to a renormalization coefficient for the single-particle Green's function. Here the definition is different, but the physics is similar, since Z again turns out to be another type of renormalization function. Now it is easy to solve the coupled equations (10.171) by writing

$$ip - \xi_p - S(p) = ipZ(p) - \bar{\xi}_p, \qquad \bar{\xi}_p = \xi_p + S_e(p) \tag{10.174}$$

$$\mathscr{G}(p) = \frac{ipZ(p) + \bar{\xi}_p}{[ipZ(p)]^2 - [\bar{\xi}_p^2 + W(p)^2]} \tag{10.175}$$

$$\mathscr{F}(p) = \frac{-W(p)}{[ipZ(p)]^2 - [\bar{\xi}_p^2 + W(p)^2]} \tag{10.176}$$

These equations are similar to those for the BCS theory (10.58). The present theory requires that the functions $W(p)$, $Z(p)$, and $S_e(p)$ are found self-consistently.

Eliashberg examined the retarded function $\mathrm{Re}[S_e(\mathbf{p}, \omega)]$ in great detail. He showed that it was a constant, independent of $(\mathbf{p}, \omega)$, which therefore just renormalizes the chemical potential. This conclusion should not be surprising. The earlier investigation in Sec. 7.4 of the electron–phonon properties of normal metals showed that $\mathrm{Re}[\Sigma(\mathbf{p}, \omega)]$ was an asymmetric function of frequency, so that the symmetric part of the self-energy is small. This conclusion does not change in the superconductor. Neglect S_e.

The other functions of interest $Z(p)$ and $W(p)$ are both functions of $(\mathbf{p}, ip)$. The wave vector dependence, for s-wave superconductors, is unimportant, since everything happens at the Fermi surface $p \approx k_F$. It is the energy dependence of these functions which provides the interesting phenomena. Actual crystalline superconductors have anisotropic Fermi surfaces, which cause the functions to vary with angle around the Fermi surface. This dependence is neglected, and isotropy is assumed. In d-wave or p-wave superconductors the angular dependence of the order parameter is also important around the Fermi surface.

For s-wave superconductors the notation is altered to suppress the dependence on wave vector. These functions are written as $S_o(ip)$, $W(ip)$, $Z(ip)$, etc. The main dependence of $\mathscr{G}(p)$ and $\mathscr{F}(p)$ on $\mathbf{p}$ is through the factor $\bar{\xi}_p \approx \xi_p$. All the wave vector dependence enters through the integrals of the Green's functions:

$$\int_{-\infty}^{\infty} d\xi_p \mathscr{G}(\mathbf{p}, ip_n) = -\int_{-\infty}^{\infty} d\xi_p \frac{ip_nZ(p) + \xi_p}{\xi_p^2 + W^2 + (p_nZ)^2} = -\frac{\pi ip_nZ}{\sqrt{p_n^2Z^2 + W^2}}$$

$$\int_{-\infty}^{\infty} d\xi_p \mathscr{F}(\mathbf{p}, ip_n) = \int_{-\infty}^{\infty} d\xi_p \frac{W(p)}{\xi_p^2 + W^2 + (p_nZ)^2} = \frac{\pi W(p)}{\sqrt{p_n^2Z^2 + W^2}}$$

These integrals are evaluated by contour integration, say by closing the contour in the upper half plane.

The earliest work on superconductivity converted these gap equations to real frequency and solved them as integral equations over the energy and frequency. That process is cumbersome, since each retarded function (W, Z) has a real and imaginary part. An alternate procedure was done by Vidberg and Serene (1977). They solved the gap equations as a function of complex frequency. The advantage is that all functions are real when expressed as a function of complex frequency. The solution to the self-consistent gap equation is much easier using only real functions. The hard part of the numerical work is then the analytical continuation of the numerical results to the real frequency axis. Since $p_n = (2n+1)\pi k_B T$, the relevent information is $(2n+1)$. Label these functions as $W(ip_n) \equiv W(2n+1)$, $S(ip_n) \equiv S(2n+1)$, etc. The end result of the computer solution is a pair of vectors representing these two functions. The best way to convert them to the axis of real frequency is to use Padé approximants (Baker, 1975). Vidberg and Serene showed this method gives excellent results.

The goal is to evaluate the gap Eqns. (10.166) using the Green's functions in (10.176). These integral equations will be reduced to a single independent variable, which is the imaginary frequency. The first step is to do all the wave vector integrals, whose variables are

$$\int d^3q = \int q^2 dq \int d\bar{\nu} \int d\phi \tag{10.178}$$

where $\bar{\nu} = \cos(\theta)$ is the angle between $\mathbf{p}$ and $\mathbf{q}$. Change the integration variable to $p' = \sqrt{p^2 + q^2 + 2pq\bar{\nu}}$, or $p'dp' = pqd\bar{\nu}$, and as a second step change $p'dp' = md\xi'$:

$$\int d^3q = \frac{m}{p}\int_0^\infty qdq \int d\xi' \int_0^{2\pi} d\phi = \frac{1}{v_p}\int d^2q \int d\xi' \tag{10.179}$$

The q integral is over the Brillouin zone; the inclusion of Umklapp processes extends the integration to infinity. The above change of variables restricts $|p - p'| \le q \le |p + p'|$. Since both (p, p') are near to k_F, the wave vector integral is restricted to $0 \le q \le 2k_F$. The $d\xi'$ integral will be limited to values of $|\xi'| < \omega_D$ at the Fermi surface, since the integral converges rapidly outside of this range. With this change of variables, the wave vector integrals are quite simple to do, at least in a formal sense. The integral over $d\xi'$ was already performed in (10.177). The remaining integration over the coordinates d^2q provides a constant μ_c for the electron–electron interaction, and the exact definition of the function $\alpha^2F(\omega)$ for the phonon term. The latter was defined in Sec. 6.4

$$\mu_c = \frac{1}{\hbar v_F}\int \frac{d^2q}{(2\pi)^3}\frac{v_q}{\varepsilon(q)} = \frac{e^2}{\pi\hbar v_F}\int_0^{2k_F}\frac{dq}{q\varepsilon(q)} \tag{10.180}$$

$$\alpha^2F(\omega) = \frac{1}{\hbar v_F}\int \frac{d^2q}{(2\pi)^3}\sum_\lambda \bar{M}_\lambda(\mathbf{q})^2\delta[\omega - \omega_\lambda(\mathbf{q})] \tag{10.181}$$

After performing these wave vector integrations, the self-consistent equations (10.166) for the two self-energy functions have been reduced to manageable proportions:

$$\begin{aligned} S(ip_n) &= -\frac{i\pi}{\beta}\sum_{iq_m}\frac{(p_n + q_m)Z(ip + iq)}{\sqrt{(p_n + q_m)^2Z^2 + W(ip + iq)^2}}\left(2\int \omega d\omega \frac{\alpha^2F(\omega)}{\omega^2 + q_m^2} - \mu_c\right) \\ W(ip_n) &= \frac{\pi}{\beta}\sum_{iq_m}\frac{W(ip + iq)}{\sqrt{(p_n + q_m)^2Z^2 + W(ip + iq)^2}}\left(2\int \omega d\omega \frac{\alpha^2F(\omega)}{\omega^2 + q_m^2} - \mu_c\right) \end{aligned} \tag{10.182}$$

The Coulomb term μ_c is a dimensionless constant, which acts to reduce the gap function.

The formulas (10.182) for $S(ip_n)$ and $W(ip_n)$ can be simplified. First change the notation to $S(2n+1)$, $W(2n+1)$. Secondly, introduce the gap function $\Delta(2n+1)$ and phonon interaction $V(m)$

$$\Delta(2n+1) = \frac{W(2n+1)}{Z(2n+1)} \tag{10.183}$$

$$V(m) = 2\int \omega d\omega \frac{\alpha^2 F(\omega)}{\omega^2 + q_m^2} \tag{10.184}$$

Note that $V(0) = \lambda$, where here λ is the dimensionless strength of the electron–phonon interaction. Also note that $V(m) < \lambda$ for $m \neq 0$. The coupled gap equations are, after changing summation variable to $k_m = p_n + q_m$

$$S(2n+1) = -\frac{i\pi}{\beta}\sum_m \Lambda(2m+1)[V(n-m) - \mu_c] \tag{10.185}$$

$$\Lambda(2m+1) = \frac{k_m}{\sqrt{(k_m)^2 + \Delta(2m+1)^2}} \tag{10.186}$$

$$Z(2n+1)\Delta(2n+1) = \frac{\pi}{\beta}\sum_m \Phi(2m+1)[V(n-m) - \mu_c] \tag{10.187}$$

$$\Phi(2m+1) = \frac{\Delta(2m+1)}{\sqrt{(k_m)^2 + \Delta(2m+1)^2}} \tag{10.188}$$

Examine the parity of these functions. The phonon interaction $V(-m) = V(m)$ is an even function of its argument. The discrete frequency is odd: $k_n = (2n+1)\pi k_B T$ changes sign if $(2n+1) \to -(2n+1)$. This feature makes $\Lambda(2n+1)$ an odd function ($\Lambda(-2n-1) = -\Lambda(2n+1)$), which also makes $S(2n+1)$ an odd function. This term is identified as $S(2n+1) = ip_n S_0(ip_n)$, so the factor of Z is

$$Z(2n+1) = 1 + \frac{1}{2n+1}\sum_m \Lambda(2m+1)[V(n-m) - \mu_c] \tag{10.189}$$

$$= 1 + \frac{1}{2n+1}\sum_m \Lambda(2m+1)V(n-m) \tag{10.190}$$

where the term in μ_c vanishes when one sums over positive and negative values of m due to the oddness of Λ. The above formula (10.190) can be used to show that $Z(2n+1)$ is an even function of its argument (i.e., change $m = -m' - 1, n = -n' - 1$).

The remaining issue is the symmetry of the gap function $\Delta(2n+1)$ as defined by Eq. (10.187). Since Z and V are both even functions, then the gap function can be either pure even or pure odd. A gap function which is odd in frequency has the feature that it must vanish at zero frequency $\Delta(ip_n = 0) = 0$. In this case it is not superconducting. The physical solution has the gap an even function of frequency. The two equations to be solved are recollected here:

$$\begin{aligned} Z(2n+1) &= 1 + \frac{1}{2n+1}\sum_m \Lambda(2m+1)V(n-m) \\ Z(2n+1)\Delta(2n+1) &= \pi k_B T \sum_m \Phi(2m+1)[V(n-m) - \mu_c] \end{aligned} \tag{10.191}$$

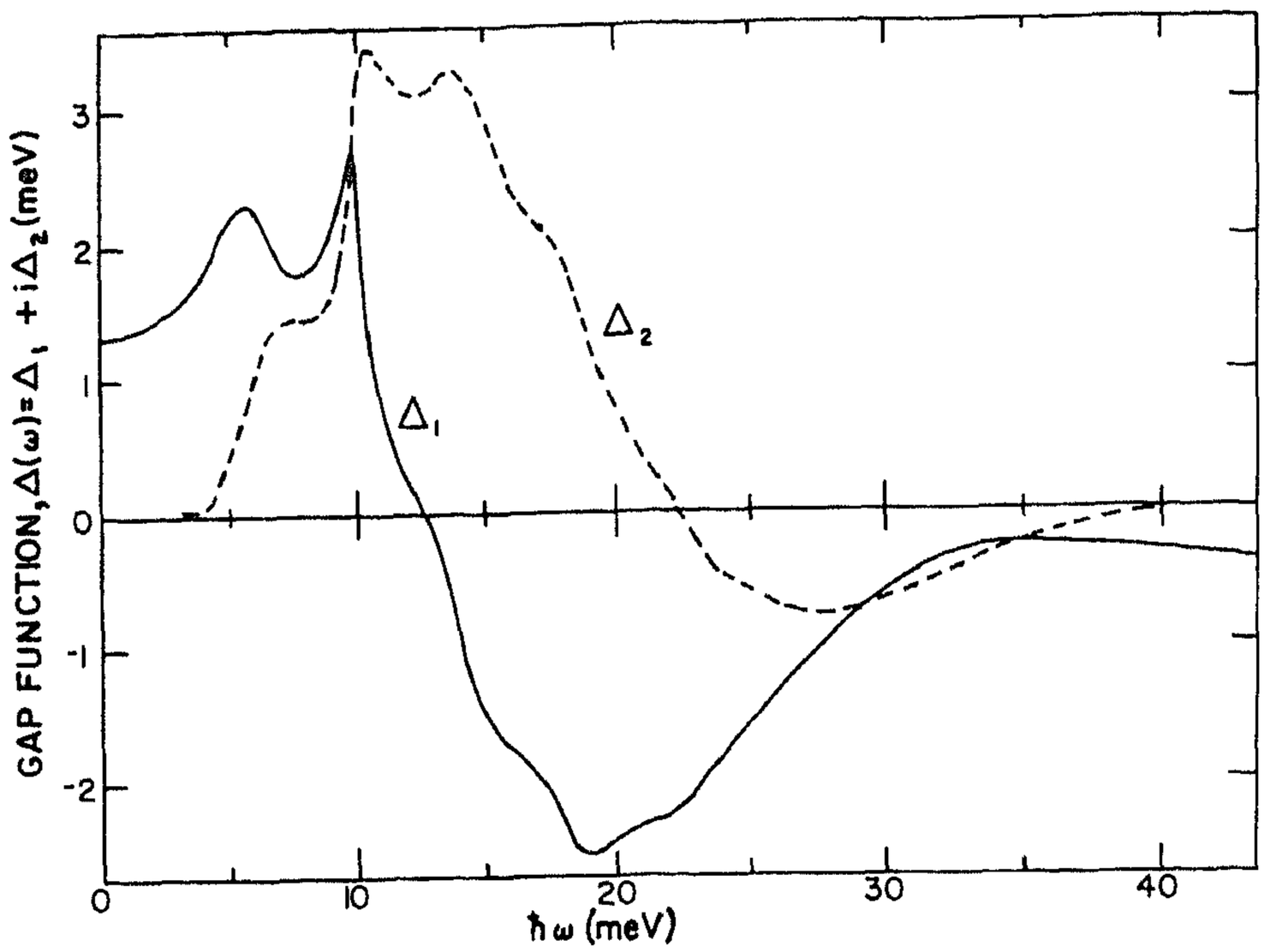


FIGURE 10.21 The real and imaginary parts Δ_1 and Δ_2 of the complex gap function as a function of energy ω for lead at $T = 0$. *Source:* Shaw and Swihart (1968) (used with permission).

These two equations must be solved self-consistently for the two real functions $\Delta(2n+1)$ and $Z(2n+1)$. The inputs for the calculation are the kernels $V(m)$, which depends on the phonon properties through α^2F, and the Coulomb factor μ_c. Values of the Coulomb constant μ_c for actual superconductors are about 0.1–0.2. The temperature T enters explicitly in the prefactor T in the equation for Δ. It also enters into the function k_n which is found in λ, Φ, and V.

The first solution to the gap function, using real frequency, was by Shaw and Swihart (1968) for lead at $T = 0$ as shown in Fig. 10.21. They used the α^2F of McMillan and Rowell (1965), which is shown in Fig. 7.13(a). The real part of the gap function $\mathrm{Re}[\Delta(\omega = 0)] \equiv \Delta_1$ equals the value 1.34 meV found from electron tunneling. The imaginary part of the gap function $\mathrm{Im}[\Delta(\omega)] \equiv \Delta_2$ is zero for $\omega < \Delta_0$. Both real and imaginary parts of $\Delta(\omega)$ are obviously strong functions of ω. Shaw and Swihart also calculated the infrared absorption and showed excellent agreement with the results of Palmer and Tinkham. The electron tunneling data for α^2F, along with the strong coupling theory, provides an excellent description of other experiments such as infrared absorption. The later calculations of Vidberg and Serene (1977) using the method of imaginary frequency and Padé approximants, obtained the same result for Pb.

10.4. TRANSITION TEMPERATURE

The transition temperature of a superconductor is called T_c. It is one of the easier quantities to calculate. A simple method was suggested by Owen and Scalapino (1971). It has been adopted widely, and is described here. The trick is to retain the imaginary frequency throughout the calculation. One advantage of using imaginary frequency is that all of the quantities are real functions, which simplifies the computation. A second advantage of the method is that it is relatively easy to include embellishments such as the influence on T_c of impurity scattering, electron–electron interactions, etc. Here a calculation is presented for the

simple case of an isotropic s-wave superconductor. The extension to more complex systems is described by Allen and Mitrović (1982).

For temperatures larger than T_c the superconducting energy gap is zero. At temperatures equal to T_c the gap starts to have nonzero value. The transition temperature is the place where the energy gap is infinitesimally small. The value of T_c is found by setting $\Delta = 0$ in all of the denominators of (10.191)

$$Z(2n+1) = 1 + \frac{1}{2n+1}\sum_m \operatorname{sgn}(2m+1)V(n-m) \tag{10.192}$$

$$Z(2n+1)\Delta(2n+1) = \sum_m \frac{\Delta(2m+1)}{|2m+1|}[V(n-m) - \mu_c] \tag{10.193}$$

where $\operatorname{sgn}(x)$ has the sign of x. The equation for $Z(2n+1)$ no longer depends upon the gap $\Delta(2n+1)$. It can be evaluated at the start of the calculation. Recall that Z comes from the self-energy of the electron. At T_c the gap is zero so one can use the self-energy of the normal state.

The function $Z(2n+1)$ now has a simple form. This fact is demonstrated by evaluating it for small values of $(2n+1)$. In the summation over m on the right, the terms are presented in two brackets: the first is when $(2m+1)$ has positive values as $m = 0, 1, 2$, etc. The second is when $2m+1$ has negative values as $m = -1, -2, -3$ etc. Note that $V(-m) = V(m)$ is symmetric. Examine the four cases that $n = 0, -1, 1, -2$

$$\begin{aligned}
Z(1) &= 1 + \{[V(0) + V(1) + V(2) + \cdots] - [V(1) + V(2) + \cdots]\} \\
&= 1 + V(0) = 1 + \lambda \\
Z(-1) &= 1 + (-1)\{[V(1) + V(2) + V(3) + \cdots] - [V(0) + V(1) + V(2) + \cdots]\} \\
&= 1 + V(0) = Z(1) \\
Z(3) &= 1 + \tfrac{1}{3}\{[V(1) + V(0) + V(1) + V(2) + \cdots] - [V(2) + V(3) + \cdots]\} \\
Z(3) &= 1 + \tfrac{1}{3}[V(0) + 2V(1)] \\
Z(-3) &= 1 + \tfrac{1}{3}[V(0) + 2V(1)] = Z(3) \\
Z(2n+1) &= 1 + \frac{1}{|2n+1|}\left[V(0) + 2\sum_{l=1}^{n} V(l)\right] = Z(-2n-1)
\end{aligned}$$

The renormalization factor $Z(2n+1)$ is an series with $2n+1$ terms. Note that $V(0) = \lambda$, where λ is the dimensionless electron–phonon coupling constant. The renormalization function is an even function of imaginary frequency.

Now examine the gap equation (10.193). The gap $\Delta(2n+1)$ is infinitesimally small. The gap equation is a linear matrix equation. The unknown quantity $\Delta(2n+1)$ is a vector. The gap equation is actually an eigenvalue equation. The solution to this equation is obtained by setting the determinant of the matrix equal to zero

$$\det|M_{nm}| = 0 \tag{10.194}$$

$$M_{nm} = \delta_{nm}Z(2n+1) - \frac{V(n-m) - \mu_c}{|2m+1|} \tag{10.195}$$

This "eigenvalue" equation is unusual, since there is no "eigenvalue". The temperature T is contained in the factor $V(m)$ which enters into Z and M_{nm}. This temperature is varied until one finds the highest value at which the determinant vanishes. This temperature is then T_c. In practice the determinant does not have to have a large dimension. For $\lambda > 1$ then its

dimension need only be about $10(-9 \leq 2n+1 \leq +9)$ for accurate answers. This dimension of determinent is easy for present computers. The solution is a symmetric function of frequency, in that $\Delta(-(2n+1)) = \Delta(2n+1)$. Knowing this fact, the dimension of the matrices can be reduced to half.

The Einstein model provides a simple example. Then all of the phonons have the same frequency ω_0 and $\alpha^2 F(\omega) = \omega_0 \lambda \delta(\omega - \omega_0)/2$. In this case it is useful to introduce the dimensionless constant $b = 2\pi k_B T/(\hbar\omega_0)$. The gap equation becomes a search for the largest value of b which satisfies the equation

$$V(m) = \frac{\lambda}{1 + b^2 m^2} \tag{10.196}$$

$$\Delta(2n+1)Z(2n+1) = \lambda \sum_m \frac{\Delta(2n+2m+1)}{|2n+2m+1|[1+b^2m^2]} \tag{10.197}$$

$$Z(2n+1) = 1 + \frac{\lambda}{|2n+1|}\left[1 + 2\sum_{l=1}^{n} \frac{1}{1+b^2 l^2}\right] \tag{10.198}$$

The gap equation (10.197) for the Einstein model of phonons is easily solved to give a graph of b vs. λ, which is shown in Fig. 10.22. Above $\lambda > 1$ the curve rises slowly with increasing coupling strength λ. In fact the dependence goes as $b \to \sqrt{\lambda}$. There is no saturation of $b = 2\pi k_B T_c/\hbar\omega_0$, which means no saturation of T_c. The higher is λ then the higher is T_c.

The asymtotic limit of large λ can be found in a simple way. Assuming that b becomes very large, then $V(m)$ becomes increasingly small for modest values of m. In the limit of large b retain only the values $V(0) = \lambda$ and $V(1) = \lambda/(1+b^2)$. The gap equation can be solved by using only a matrix of dimension two. It is necessary to retain only the gap components $\Delta(1)$ and $\Delta(-1)$, and the renormalization parameter $Z(1) = Z(-1) = 1 + \lambda$. When $\mu_c = 0$ the equations are

$$\Delta(1)(1+\lambda) = \lambda\left\{\Delta(1) + \frac{\Delta(-1)}{1+b^2}\right\} \tag{10.199}$$

$$\Delta(-1)(1+\lambda) = \lambda\left\{\Delta(-1) + \frac{\Delta(1)}{1+b^2}\right\} \tag{10.200}$$

The matrix equation is 2×2. Setting the determinent equal to zero gives two solutions for $b(\lambda)$. The only one for real values of b is

$$b^2 = \lambda - 1 \tag{10.201}$$

$$k_B T_c = \frac{\hbar\omega_0}{2\pi}\sqrt{\lambda - 1} \tag{10.202}$$

Keep in mind that this solution is approximate and only valid in the limit that $\lambda \to \infty$. However, this solution clearly shows that T_c increase as the square root of the electron–phonon coupling constant at large values of the coupling constant. The exact asymptotic limit as $\lambda \to \infty$ is $k_B T_c \to 0.15\hbar\omega_0\sqrt{\lambda}$ (Allen and Mitrović, 1982).

The electron–electron parameter μ_c plays the same role as $V(m)$ for the electron–phonon interaction. Combining these two interactions gives a gap equation with an effective interaction of $(V - \mu_c)$. For $\lambda > \mu_c$ the interaction is attractive for small values of m but becomes negative for larger values of m. The electron–phonon interaction is attractive only over a limited range of imaginary frequency.

Figure 10.22(a) shows the solution of the gap equation for the Einstein model assuming that $\mu_c = 0.1$. The solid line has $\mu_c = 0$ compared to the dashed line with $\mu_c = 0.1$. Electron–

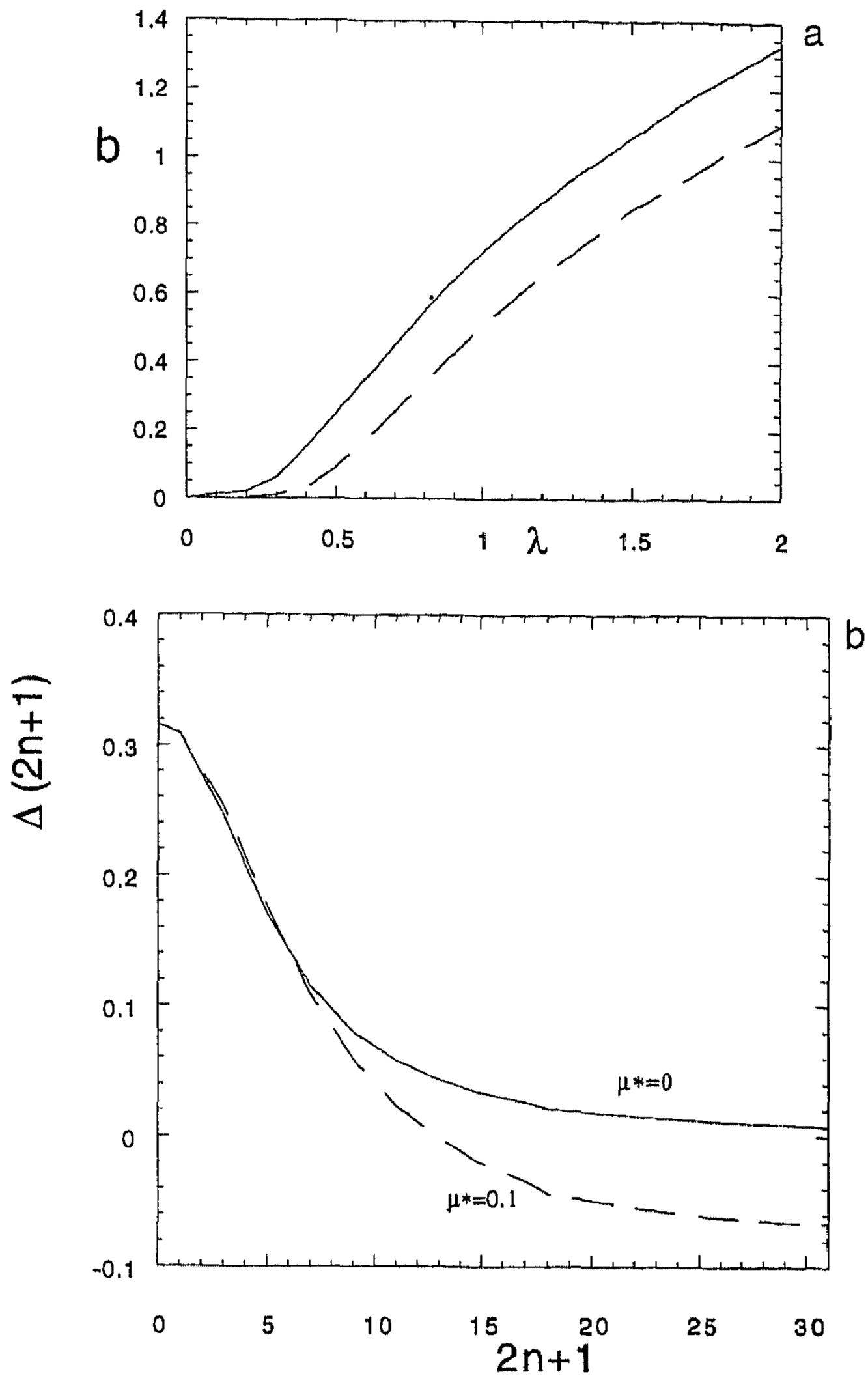


FIGURE 10.22 (a) The solution to the gap equations is given as $b(\lambda)$. The solid line has $\mu_c = 0$, while the dashed line has $\mu_c = 0.1$. (b) The eigenvectors $\Delta(2n+1)$ for the two cases.

electron interactions tend to reduce the value of $b = 2\pi k_B T_c/\hbar\omega_0$. Another interesting function are the eigenvectors $\Delta(2n+1)$. Figure 10.22(b) shows the relative values of this function for the case that $\lambda = 1$. The two curves are the cases that $\mu_c = 0.1$ and 0. Since the gap function is symmetric in $(2n+1)$ only positive values are shown. Of course, the gap function vanishes at T_c. Here we assume the temperature is approaching T_c and the gap functions are very small. In this case only their relative size matters. Note that the two curves are very similar near to small values of frequency $(2n+1)$. However, they differ markedly at large frequency. For the case of no electron–electron interaction ($\mu_c = 0$), the relative values of $\Delta(2n+1)$ get very small and vanish at large $(2n+1)$ which justifies the use of matrices of relatively small dimension in solving the gap equation.

The opposite behavior is found when $\mu_c \neq 0$. At large values of $(2n+1)$, the gap function $\Delta(2n+1)$ does not go to zero but goes to a constant. Then there is no argument for using a determinent of small dimension. The problem is that the effective interaction in the gap equation $V(m) - \mu_c$ does not vanish at large m but goes to a constant $-\mu_c$. Therefore one should use large determinents. However, there is a method of using determinents of smaller

dimension, which entails rescaling the Coulomb coefficient. This procedure is described in the review by Allen and Mitrović (1982).

Problems

1. Consider the mutual scattering of a fermion and a boson, i.e., a ^{3}He and a ^{4}He particle moving parallel in time. Use Green's functions to derive the occupation number factors which enter the scattering integral.

2. Calculate how the pole in the Cooper instability varies with the center of mass momentum **P** of the pair.

3. What is the momentum distribution $n_{\mathbf{k}}$ of electrons in the superconductor at $T = 0$?

4. The entropy per unit volume of a superconductor in the BCS model is

$$S = -\frac{2k_B}{V}\sum_{\mathbf{k}}\{n_F(E_k)\ln[n_F(E_k)] + [1 - n_F(E_k)]\ln[1 - n_F(E_k)]\}$$

Use this expression to derive the specific heat of the superconductor. Then use $\Delta(T) \approx \sqrt{T_c - T}$ to show that the specific heat is discontinuous at the critical temperature.

5. Write the Hamiltonian (10.35) in the following form:

$$H_{\text{eff}} = \sum_{\mathbf{k}\sigma}\xi_k C^\dagger_{\mathbf{k}\sigma}C_{\mathbf{k}\sigma} - \sum_{\mathbf{k}}\Delta(\mathbf{k})(C^\dagger_{\mathbf{k}\uparrow}C^\dagger_{-\mathbf{k}\downarrow} + C_{-\mathbf{k}\downarrow}C_{\mathbf{k}\uparrow})$$

$$\Delta(\mathbf{k}) = -\sum_{\mathbf{q}} V(q)\langle C_{-\mathbf{k}-\mathbf{q}\downarrow}C_{\mathbf{k}+\mathbf{q}\uparrow}\rangle$$

Solve the effective Hamiltonian H_{eff} by a canonical transformation, and reduce it to the form

$$H_{\text{eff}} = E_0 + \sum_{\mathbf{k}} E_k(\alpha^\dagger_{\mathbf{k}}\alpha_{\mathbf{k}} + \beta^\dagger_{\mathbf{k}}\beta_{\mathbf{k}}) \tag{10.203}$$

$$C_{\mathbf{k}\uparrow} = \cos(\theta_{\mathbf{k}})\alpha_{\mathbf{k}} + \sin(\theta_{\mathbf{k}})\beta^\dagger_{\mathbf{k}} \tag{10.204}$$

$$C_{-\mathbf{k}\downarrow} = \cos(\theta)\beta_{\mathbf{k}} - \sin(\theta_{\mathbf{k}})\alpha^\dagger_{\mathbf{k}} \tag{10.205}$$

6. Calculate the Pauli spin susceptibility of the BCS superconducting state. Show that the formula at nonzero temperatures is

$$\chi = 2\mu_0^2\sum_{\mathbf{k}}(d/dE_k)n_F(E_k)$$

7. Derive the formula for the coefficient in the Josephson current $J_S(0) = (\pi\Delta/2)(\sigma_0/e)$ by doing the integral in (10.121) for $eV = 0$.

8. Evaluate $J_S(eV)$ in (10.127) at zero temperature by making the change $k_BT\sum_{ip} \to \int dp/(2\pi)$ and then the analytical continuation $i\omega \to eV$. Define $\Lambda(eV)$ at $T = 0$ by using $J_S(eV) = (\sigma_0/e)\Lambda(eV)$. Show that:

(a) $\Lambda(eV = 0) = \Delta_s K(q)$: $\quad q = \left(1 - \dfrac{\Delta_S^2}{\Delta_L^2}\right)^{1/2}$

where Δ_S and Δ_L are the smaller and larger of the two energy gaps.

(b) When $\Delta_L = \Delta_S = \Delta$, then $\Lambda(eV) = \Delta K(eV/2\Delta)$.

9. Use strong coupling theory to derive a formula for the coefficient $J_S(eV)$ for the Josephson tunneling.

10. Derive the strong coupling theory formula for the infrared absorption of a dirty superconductor.

11. At large values of λ the gap equation for $\mu_c = 0$ goes as $b \to \alpha_0\sqrt{\lambda}$ where the best numerical solution gives $\alpha_0 = 0.1481$ (Mahan, 1997). Our first estimate solved the gap equation retaining only $\Delta(\pm 1)$, and gave $\alpha_0 = 1$. Make a better estimate of this constant by solving the gap equation while retaining $\Delta(\pm 1)$, $\Delta(\pm 3)$. (Hint: one cannot solve for b analytically. But letting $\xi = \lambda/b^2$ so that $V(m) = \xi/m^2$ gives a simple equation for ξ.)

12. How does the electron–electron parameter μ_c change the result in (10.202)?

13. For impurity scattering, take the self-energy of an electron to be given in terms of a T matrix

$$\Sigma(\mathbf{p}, ip_n) = n_i \int \frac{d^3q}{(2\pi)^3} |T_{\mathbf{p},\mathbf{p}+\mathbf{q}}|^2 \frac{1}{ip_n - \varepsilon_{\mathbf{p}+\mathbf{q}}} \tag{10.206}$$

This interaction does not change the energy of the electron, but does change its wave vector. Use this function to calculate a parameter

$$\gamma_i = \beta \int \frac{d^3q}{(2\pi)^3} |T_{\mathbf{p},\mathbf{p}+\mathbf{q}}|^2 \delta(\varepsilon - \xi_{\mathbf{p}+\mathbf{q}}) \tag{10.207}$$

which enters into the gap equation and also into Z. Show that these two contributions cancel in the gap equation for an isotropic superconductor. In this case, impurity scattering does not change the value of T_c. This was first shown by P.W. Anderson.

14. Use the Thomas–Fermi model for the dielectric function to evaluate μ_c for the homogeneous electron gas. Give numerical estimates for $r_s = 2, 3, 4$.

15. Repeat the above calculation for the two dimensional electron gas, where the electron dispersion is parabolic and the Fermi surface is a circle. The major change is in the dielectric screening function.

Chapter 11

Superfluids

Helium has two common isotopes, ^{3}He and ^{4}He. Each isotope can be separated and a liquid formed at low temperatures which is nearly pure ^{3}He or pure ^{4}He. Each has unusual properties and displays collective behavior of a unique character. The boson liquid ^{4}He shows a phase transition at $T_\lambda = 2.172\,\mathrm{K}$ to a superfluid state which is similar to Bose–Einstein condensation, although vastly modified by the strong interparticle interactions. Similarly, the fermion liquid ^{3}He develops a Fermi distribution at low temperature, and the particles have a superfluid transition which is similar to the superconducting transition in a metal. Of course, now it is occurring in a liquid, of electrically neutral atoms, so there is no Meissner effect, but there is pairing. However, it also has a unique character, since the atoms avoid the usual singlet pairing common to metals and instead pair with the spins aligned parallel. The triplet pairing, in turn, leads to many new phenomena and a richer phase diagram, which was discovered by Osheroff *et al.* (1972). The third superfluid in this chapter is the two-dimensional electron gas under strong magnetic field, which shows the quantum Hall effect and the fractional quantum Hall effect. It is also a highly correlated fluid.

Quantum fluids is a subject where Green's functions are often abandoned in favor of other techniques. There are some calculations for which other techniques are better and for which the Green's function method gives awful results. One example is the discussion of the ground state properties of quantum liquids. Green's function techniques have been unsuccessful in the study of the ground state properties of classical fluids, and other techniques are now used for these high-density, strongly interacting, and highly correlated systems (see Percus, 1964). It is understandable that the ground state of quantum liquids is similarly difficult to treat with Green's functions. The description of the ground state properties of the system will require techniques beyond those introduced here, and often these results only summarize. However, when discussing the excitation spectra of the liquid, which are crucial for their superfluid properties, familiar techniques are used, such as Green's functions or perturbation theory.

11.1. LIQUID ^{4}He

The first discussion of ^{4}He uses pairing theory. The important approximation in pairing theory is that the particles' interactions can be approximated as a summation of pairwise events. Of course, the potentials between particles are a summation of pair interactions:

$\sum V(\mathbf{r}_i - \mathbf{r}_j)$. But pairing theory makes a much stronger assumption: when one particle interacts with another, it is not simultaneously interacting with other particles. Pairing theory would be a good approximation in a gas of low density, where the particles collide occasionally, and most collision events are binary. But this assumption is not obviously valid in a liquid. In fact, it is now known that the assumption is terrible and that pairing theory is a very bad approximation when applied to liquid helium. It makes numerical predictions of measurable quantities, such as the speed of sound in ^{4}He, which are a factor of 10 higher than experiment. One should not regard the pairing theory as a serious theory of liquid helium.

There are still several reasons for introducing pairing theory. Historically it played an important role in the history of many-particle physics. The major ideas were introduced by Bogoliubov (1947), who proposed the idea that the particles interacted in pairs. Although this theory was unsuccessful when applied to helium, virtually the same theory works for superconductivity. The BCS theory is just the Bogoliubov theory, changing the operators from bosons to fermions. Of course, there it works since the particles do have pairwise interactions and the interactions are weak. In superfluid helium it does not work, although much effort was expended over a period of years by many investigators before this conclusion became obvious.

Another reason for introducing pairing theory is that it contains many ideas which are qualitatively correct such as the condensate. It provides a good introduction to the subject. Later it is shown how the quantum liquid properties are actually calculated using correlated basis functions.

The starting point for this discussion is a Hamiltonian which contains the kinetic energy of the particles and also the pairwise potential $V(r)$

$$H = \sum_j \frac{p_j^2}{2m} + \tfrac{1}{2}\sum_{i\neq j} V(\mathbf{r}_i - \mathbf{r}_j) \tag{11.1}$$

The helium particles are treated as spherically symmetric objects, and the electronic excitations are ignored. This approximation is accurate, since the atoms have a kinetic energy of roughly 15 K, using temperature units, while the electronic excitations have a minimum of 20 eV. The ^{4}He atom is considered to be entirely structureless and is represented as a single-boson particle with a potential $V(r)$ when interacting with other similar bosons. The ^{3}He atom has additional degrees of freedom associated with its nuclear spin. The spin is an important property, since it is what makes ^{3}He have its fermion character. The helium–helium potential which was used most frequently in the early days of physics was a Lennard-Jones potential,

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{11.2}$$

$$\varepsilon = 1.484 \times 10^{-22}\ \mathrm{J} \tag{11.3}$$

$$\sigma = 2.648\ \text{Å} \tag{11.4}$$

which is often called a 6–12 potential. The parameter $\varepsilon \approx 10$ K is the maximum well depth, and σ is the hard core radius, where the potential rises steeply. These parameters were obtained by fitting various experimental results to this potential function (see Hirschfelder *et al.*, 1954). It should be appreciated that this potential form was chosen for its mathematical simplicity and not because it was a good approximation to the actual potential shape. Certainly the long-range attractive potential behaves as r^{-6} from van der Waals forces, but the repulsive part is probably not r^{-12}. Recent efforts to deduce this potential function have produced better versions.

In the pairing theory the Hamiltonian is written $H = H_0 + V$, where H_0 is the kinetic energy term. All operators are expanded in the basis set of H_0, which are plane-wave states. Rewrite the Hamiltonian (11.1) as

$$H = \sum_{\mathbf{p}} \xi_p C_{\mathbf{p}}^\dagger C_{\mathbf{p}} + \frac{1}{2v} \sum_{\mathbf{qkk'}} V(\mathbf{q}) C_{\mathbf{k}-\mathbf{q}}^\dagger C_{\mathbf{k}'+\mathbf{q}}^\dagger C_{\mathbf{k}'} C_{\mathbf{k}} \tag{11.5}$$

$$\xi_p = \frac{p^2}{2m} - \mu \tag{11.6}$$

An additional spin index should be added when describing ^{3}He. The effective interaction $V(\mathbf{q})$ is the Fourier transform of the interparticle potential $V(r)$. One immediate problem is that the Fourier transform may not exist. Certainly there is no Fourier transform for a Lennard-Jones potential. It is shown later that the potential can be replaced by a T matrix, which always exists regardless of the potential strength or shape. This divergence is formally eliminated until the T matrix is introduced by assuming the divergence is cut off by a parameter g and by later letting $g \to \infty$. The nonexistence of $V(\mathbf{q})$ is not really a serious problem, since the replacement of scattering properties by a T matrix is formally exact for binary collisions. This theory would give a good description of a dilute gas.

11.1.1. Hartree and Exchange

Liquid helium is not a weakly interacting system, but it is treated as such in pairing theory. It is assumed, for the present discussion, that the potential terms are weak. The self-energy is found from the first terms which occur in the S-matrix expansion. The first two contributions are identical to the terms found for the homogeneous electron gas, which has the same formal kind of Hamiltonian as (11.5). They are the Hartree and exchange energies:

$$\Sigma(\mathbf{k}) = n_0 V(\mathbf{q} = 0) \pm \frac{1}{v} \sum_{\mathbf{q}} n(\xi_{\mathbf{p}+\mathbf{q}}) V(\mathbf{q}) \tag{11.7}$$

The $\pm$ signs refer to ^{4}He(+) and ^{3}He(−), respectively. The ^{3}He case is identical to that of the electron gas, where the fermion nature makes the exchange energy have the opposite sign as the Hartree energy. The fermions of like spin wish to avoid each other. For ^{4}He, its boson character makes the opposite occur: the quantum nature of the particles tends to make them prefer to collect at the same spot, so that the exchange energy is positive. For a system with only two spinless bosons, the two-particle wave function must be symmetric under an interchange of coordinates, which can be achieved with the choice

$$\Psi_{\mathbf{k}_1\mathbf{k}_2}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [\phi_{\mathbf{k}_1}(\mathbf{r}_1)\phi_{\mathbf{k}_2}(\mathbf{r}_2) + \phi_{\mathbf{k}_1}(\mathbf{r}_2)\phi_{\mathbf{k}_2}(\mathbf{r}_1)] \tag{11.8}$$

$$\phi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{v}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{11.9}$$

The expectation value of the potential energy for this wave function is

$$\langle V\rangle = \sum_{\mathbf{k}_1\mathbf{k}_2} n(\xi_{\mathbf{k}_1})n(\xi_{\mathbf{k}_2}) \int d^3r_1 d^3r_2 |\Psi_{\mathbf{k}_1\mathbf{k}_2}(\mathbf{r}_1, \mathbf{r}_2)|^2 V(\mathbf{r}_1 - \mathbf{r}_2) \tag{11.10}$$

$$= \frac{1}{v^2}\sum_{\mathbf{k}_1\mathbf{k}_2} n(\xi_{\mathbf{k}_1})n(\xi_{\mathbf{k}_2}) \int d^3r_1 d^3r_2 [1 + e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot(\mathbf{r}_1-\mathbf{r}_2)}] V(\mathbf{r}_1 - \mathbf{r}_2)$$

$$= \frac{1}{v}\sum_{\mathbf{k}_1\mathbf{k}_2} n(\xi_{\mathbf{k}_1})n(\xi_{\mathbf{k}_2})[V(\mathbf{q}=0) + V(\mathbf{k}_1 - \mathbf{k}_2)] \tag{11.11}$$

In the integral in the middle equation, the integration variables were changed to center of mass coordinates $\mathbf{R} = (\mathbf{r}_1 + \mathbf{r}_2)/2$ and $\mathbf{r} = (\mathbf{r}_1 - \mathbf{r}_2)$, and $\int d^3R = v$. The remaining integral over d^3r produces $V(\mathbf{q} = 0)$ in the first term and $V(\mathbf{k}_1 - \mathbf{k}_2)$ in the second. The expression for $\langle V\rangle$ has two terms, which are the Hartree and exchange energies. They have a relative plus sign, whose origin is the relative plus sign between the two terms in the two-particle wave function (11.8). The two-particle fermion wave function, when the spins are parallel, has a relative minus sign, which causes the negative sign in front of the exchange energy. The occupation factors $n(\xi_k)$ are either n_B for ^{4}He or n_F for ^{3}He particles. The subscript is omitted in the formulas, since they can apply to either case equally. The diagrams for Hartree and exchange are shown in Figs. 11.1(a) and 11.1(b).

The next step is to improve the Hartree and exchange energies by replacing the potential $V(\mathbf{q})$ by the equivalent T-matrix result. The Hartree energy is the potential energy of a particle in $\mathbf{k}_1$ from the interaction with the other particles in the state $\mathbf{k}_2$ by the $\mathbf{q} = 0$ interaction $V(\mathbf{q} = 0)$. This interaction can be viewed as a process whereby the two particles interact by mutually scattering. Go to center of mass coordinates: $\mathbf{k} = (\mathbf{k}_1 - \mathbf{k}_2)/2$, $\mathbf{K} = (\mathbf{k}_1 + \mathbf{k}_2)$. The mutual scattering of two particles, which interact by central forces, does not alter the center of mass wave vector $\mathbf{K}$. It is unaffected by the scattering process, so that the scattering can be viewed in a rest frame where $\mathbf{K} = 0$. This scattering is shown in Fig. 11.1(c). The two particles approach with $\mathbf{k}$ and $-\mathbf{k}$, scatter with a $\mathbf{q} = 0$ interaction, and depart with their same wave vectors. In the center of mass transformation, the labeling convention is that the particle

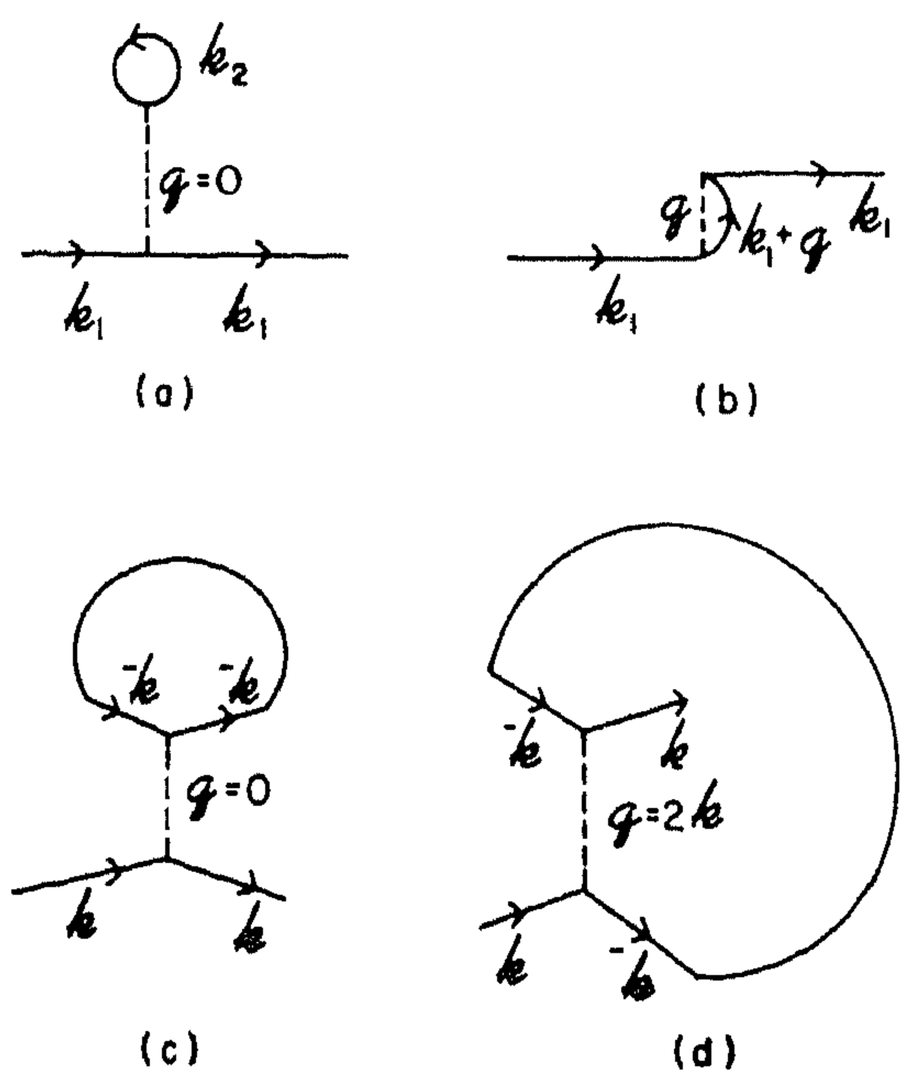


FIGURE 11.1

$\mathbf{k}$ is the one whose self-energy is being calculated, while $-\mathbf{k}$ is the other particle from which the first is scattering. The particle line of $-\mathbf{k}$ is part of a closed loop.

The scattering of two particles is described by a T matrix $T_{\mathbf{k}\mathbf{k}'}$, which was introduced in Sec. 4.1. The choice of T matrix, rather than reaction matrix, is based on wanting retarded self-energies, since the T matrix has the proper dependence on $i\delta$ in the energy denominators. Recall that the on-shell T matrix $T_{\mathbf{k}\mathbf{k}'}$ applies only for elastic scattering, when the particle energy was $\varepsilon = \xi_{\mathbf{k}} = \xi_{\mathbf{k}'}$. The kinetic energy $\xi_{\mathbf{k}}$ is conserved in relative coordinates, so this requirement is satisfied. The retarded self-energy for particles of energy $\xi_{\mathbf{k}}$ in the Hartree approximation should be

$$\Sigma_H(\mathbf{k}_1, \xi_{\mathbf{k}_1}) = \frac{1}{v}\sum_{\mathbf{k}_2} n(\xi_{\mathbf{k}_2}) T_{(\mathbf{k}_1-\mathbf{k}_2)/2,(\mathbf{k}_1-\mathbf{k}_2)/2} \tag{11.12}$$

In the scattering from other particles, whose wave vector is $\mathbf{k}_2$, the scattering rate depends on the difference $\mathbf{k} = (\mathbf{k}_1 - \mathbf{k}_2)/2$ rather than on $\mathbf{q} = 0$. The forward scattering with $\mathbf{k}' = \mathbf{k}$ is easy to obtain from Sec. 4.1,

$$T_{\mathbf{k}\mathbf{k}} = -\frac{4\pi}{2\mu k}\sum_l (2l+1)e^{i\delta_l(k)}\sin[\delta_l(k)] \tag{11.13}$$

where μ is the reduced mass. By changing integration to $\mathbf{k} = (\mathbf{k}_1 - \mathbf{k}_2)/2$ and $d^3k_2 = 8d^3k$, the Hartree energy is

$$\Sigma_H(\mathbf{k}_1, \xi_{\mathbf{k}_1}) = -\frac{16\pi}{\mu}\int \frac{d^3k}{(2\pi)^3} n(\xi_{\mathbf{k}_1-2\mathbf{k}})\frac{1}{k}\sum_l (2l+1)e^{i\delta_l(k)}\sin[\delta_l(k)]$$

The same considerations should now be applied to the scattering term for the exchange energy. Examine the scattering in the center of mass coordinates, where the incoming particles have wave vectors $\mathbf{k}$ and $-\mathbf{k}$. This diagram is shown in Fig. 11.1(d). Here the two particles exchange momentum $2\mathbf{k}$, so the final states are still $\mathbf{k}$ and $-\mathbf{k}$. The solid line indicates the particle lines which are connected to make the self-energy diagram for exchange. This term has a T-matrix expression:

$$\begin{aligned}\Sigma_x(\mathbf{k}_1, \xi_{\mathbf{k}_1}) &= \frac{1}{v}\sum_{\mathbf{k}_2} n(\xi_{\mathbf{k}_2}) T_{(\mathbf{k}_1-\mathbf{k}_2)/2,(\mathbf{k}_2-\mathbf{k}_1)/2} \\ &= -\frac{16\pi}{\mu}\int \frac{d^3k}{(2\pi)^3} n(\xi_{\mathbf{k}_1-2\mathbf{k}})\frac{1}{k}\sum_l (2l+1)(-1)^l e^{i\delta_l(k)}\sin[\delta_l(k)]\end{aligned} \tag{11.14}$$

When making the expansion in angular momentum states l, then $P_l(-1) = (-1)^l$. The two self-energy terms, Hartree and exchange, are either added or subtracted, respectively, for bosons and fermions. These two self-energy expressions are added for ^{4}He:

$$\Sigma(\mathbf{k}, \xi_{\mathbf{k}}) = -\frac{4}{\mu\pi^2}\int_0^\infty k'dk' \sum_{\text{even } l} (2l+1)e^{i\delta_l(k')}\sin[\delta_l(k')]\int d\Omega_{\mathbf{k}'} n(\xi_{\mathbf{k}-2\mathbf{k}'})$$

For ^{4}He, the two terms combine to eliminate angular momentum components l which are odd. Only the phase shifts $\delta_l(k)$ for l even need to be computed, and the others are irrelevant.

For ^{3}He, there is exchange scattering between two particles only when the spins are parallel, and then the scattering rate has no dependence upon the angular momentum terms with l even but contains only terms with l being odd. Since there is no exchange scattering between particles with antiparallel spin, the energy of a particle contains an average over

events with spin parallel and antiparallel. For the self-energy for a particle in ^{3}He then $\Sigma(\mathbf{k}) = 2\Sigma_H - \Sigma_x$.

11.1.2. Bogoliubov Theory of ^{4}He

The Bogoliubov (1947) theory describes Bose–Einstein condensation in a weakly interacting system. It is basically a form of pairing theory for the interacting bosons. The Hamiltonian (11.5) for a system of interacting bosons is expressed in the plane-wave representation. The main feature of the theory is the assumption that a Bose–Einstein condensation occurs at a nonzero temperature T_λ. The condensation occurs because the chemical potential μ goes to zero, and there is a macroscopic occupation of the zero-momentum state. The collection of particles in the zero-momentum state is called the *condensate.*

The ^{4}He particles are treated as bosons. Because the number of particles is fixed, the number operator

$$n_B(\omega_\mathbf{k}) = \frac{1}{e^{\beta(\omega_\mathbf{k}-\mu)} - 1} \tag{11.15}$$

$$N = \int \frac{d^3k}{(2\pi)^3} n_B(\omega_\mathbf{k}) \tag{11.16}$$

contains a chemical potential μ. It is usually temperature dependent and varies to satisfy the implied definition in (11.16) that the integral over the occupation number gives the total number of particles N. The particle energy $\omega_\mathbf{k}$ is yet to be determined, self-consistently, but it is positive and the zero of energy is set at $\omega_{\mathbf{k}=0} = 0$. With these conventions, the chemical potential for most boson systems is negative, so that the energy denominator in $n_B(\omega_\mathbf{k})$ is never zero because $\omega_\mathbf{k} > \mu$. However, in Bose–Einstein condensation, the chemical potential μ vanishes, as illustrated in Fig. 11.2. The occupation number $n_B(\omega_\mathbf{k})$ diverges as $\mathbf{k} \to 0$, and the number of particles in the $\mathbf{k} = 0$ states is not well defined by this limit. Instead, assume that the number of particles in the state $\mathbf{k} = 0$ is some number N_0 which is a finite fraction of the particle number N. Indeed, for a weakly interacting system, the condensate fraction $f_0 = N_0/N$ approachs a number near unity as the temperature is lowered below T_λ. For $T < T_\lambda$, the particle number operator is expressed as

$$n_B(\omega_\mathbf{k}) = N_0\delta(\mathbf{k}) + \frac{1}{e^{\beta\omega_\mathbf{k}} - 1} \tag{11.17}$$

The delta function at $\mathbf{k} = 0$ gives the number of particles in the condensate, and the second term applies only for $\mathbf{k} \neq 0$. One of the goals of the theory is to calculate this fraction f_0 self-consistently.

The creation and destruction operators $C_\mathbf{k}^\dagger$ and $C_\mathbf{k}$ must have special properties for $\mathbf{k} = 0$. Normally, for bosons, when there are $n_\mathbf{k}$ particles in the state $\mathbf{k}$, the creation operator

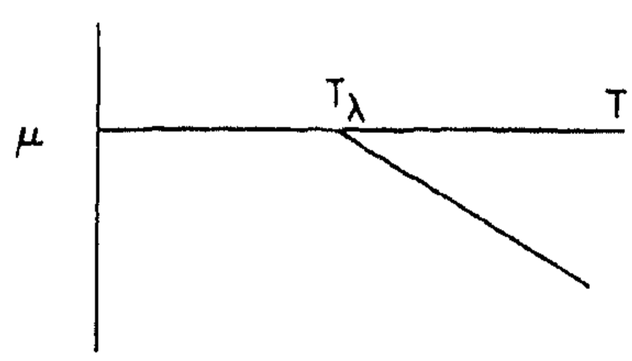


FIGURE 11.2 Chemical potential as a function of temperature. Note that μ becomes zero at T_λ in Bose–Einstein condensation.

gives $C_{\mathbf{k}}^{\dagger}|n_{\mathbf{k}}\rangle = \sqrt{n_{\mathbf{k}}+1}|n_{\mathbf{k}}+1\rangle$. In normal systems, the occupation numbers $n_{\mathbf{k}}$ are integers on the order of unity. For the state with $\mathbf{k}=0$, the average value of $\langle n_0\rangle = N_0 \approx 10^{23}$. Then the operations give

$$C_0^{\dagger}|N_0\rangle = \sqrt{N_0+1}|N_0+1\rangle \approx \sqrt{N_0}|N_0\rangle \tag{11.18}$$

$$C_0|N_0\rangle = \sqrt{N_0}|N_0-1\rangle \approx \sqrt{N_0}|N_0\rangle \tag{11.19}$$

where the numbers $N_0 \pm 1 \approx N_0$ when $N_0 \sim 10^{23}$. Both the creation and destruction operators have a similar effect upon the system and give the result $\sqrt{N_0}$ as an effective eigenvalue. The number N_0 is sufficiently large that the operators $C_0^{\dagger}$ and C_0 are treated as scalars whose value is $\sqrt{N_0}$.

$$C_0^{\dagger}C_0 = C_0C_0^{\dagger} = N_0 \tag{11.20}$$

The next step is to examine the Hamiltonian and to isolate all terms in which any of the C operators has $\mathbf{k}=0$. These terms are replaced by $\sqrt{N_0}$ for both $C_0^{\dagger}$ and C_0. Since $\mu = 0$, write the kinetic energy term as

$$\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} C_{\mathbf{k}}^{\dagger}C_{\mathbf{k}} = \varepsilon_0 N_0 + \sum_{\mathbf{k}}{}' \varepsilon_{\mathbf{k}} C_{\mathbf{k}}^{\dagger}C_{\mathbf{k}} = \sum_{\mathbf{k}}{}' \varepsilon_{\mathbf{k}} C_{\mathbf{k}}^{\dagger}C_{\mathbf{k}} \tag{11.21}$$

The prime on the summation means to omit the term with $\mathbf{k}=0$. The kinetic energy has no dependence on the condensate, since the particles have a zero kinetic energy in this state. The interesting terms arise from the potential energy:

$$\frac{1}{2v}\sum_{\mathbf{qkk}'} V(\mathbf{q}) C_{\mathbf{k}-\mathbf{q}}^{\dagger} C_{\mathbf{k}'+\mathbf{q}}^{\dagger} C_{\mathbf{k}'} C_{\mathbf{k}} \tag{11.22}$$

When the condensate fraction f_0 is near unity, the largest term is when $\mathbf{k}=\mathbf{k}'=\mathbf{q}=0$, which gives an energy contribution in terms of the condensate density $\rho_0 = N_0/v = f_0\rho$:

$$\frac{1}{2v}V(0)N_0^2 = \tfrac{1}{2}v\rho_0^2 V(0) \tag{11.23}$$

The next largest terms are those in which only three of the C operators has zero wave vector, but none exist. If any three particles have zero wave vectors, then so does the fourth, and that term has been counted already. The next largest terms are those in which two of the wave vectors are zero and two are not. There are six ways of pairing the four C operators in the potential energy term:

$$\begin{aligned}
&\mathbf{k}=\mathbf{k}'=0 && \tfrac{1}{2}\rho_0\Sigma_{\mathbf{q}}V(\mathbf{q})C_{\mathbf{q}}^{\dagger}C_{-\mathbf{q}}^{\dagger}\\
&\mathbf{k}=\mathbf{q}=-\mathbf{k}' && \tfrac{1}{2}\rho_0\Sigma_{\mathbf{q}}V(\mathbf{q})C_{\mathbf{q}}C_{-\mathbf{q}}\\
&\mathbf{q}=0=\mathbf{k} && \tfrac{1}{2}\rho_0 V(0)\Sigma_{\mathbf{k}'}C_{\mathbf{k}'}^{\dagger}C_{\mathbf{k}'}\\
&\mathbf{q}=0=\mathbf{k}' && \tfrac{1}{2}\rho_0 V(0)\Sigma_{\mathbf{k}}C_{\mathbf{k}}^{\dagger}C_{\mathbf{k}}\\
&\mathbf{k}=0,\mathbf{k}'=-\mathbf{q} && \tfrac{1}{2}\rho_0\Sigma_{\mathbf{q}}V(\mathbf{q})C_{\mathbf{q}}^{\dagger}C_{\mathbf{q}}\\
&\mathbf{k}'=0,\mathbf{k}=\mathbf{q} && \tfrac{1}{2}\rho_0\Sigma_{\mathbf{q}}V(\mathbf{q})C_{\mathbf{q}}^{\dagger}C_{\mathbf{q}}
\end{aligned} \tag{11.24}$$

It is customary to stop at this point and to keep only these terms in an effective Hamiltonian:

$$H_0 = \sum_{\mathbf{k}}' \left\{ \varepsilon_{\mathbf{k}} + \rho_0 V(0)] C_{\mathbf{k}}^\dagger C_{\mathbf{k}} + \tfrac{1}{2}\rho_0 V(\mathbf{k})[C_{\mathbf{k}}^\dagger + C_{-\mathbf{k}}][C_{\mathbf{k}} + C_{-\mathbf{k}}^\dagger] \right\} \tag{11.25}$$

The term $\rho_0 V(0)$ is dropped, since this constant term just causes a shift in the chemical potential, which is still zero. The effective Hamiltonian given in (11.25) can be diagonalized exactly, which is the strongest argument for using it as the effective Hamiltonian. A better theory would retain additional terms. There are terms with three C operators with nonzero wave vectors, and of course terms with four. The Hamiltonian is written as $H_0 + V$, where H_0 is given in (11.25). It contains the terms quadratic in the operators, while V is the other terms which contain three or four creation–destruction operators. The terms are separated in this manner because H_0 is a Hamiltonian which can be solved exactly. The other term V must be treated as the interaction, which is evaluated by the usual Green's function method to give additional self-energy corrections.

The particles in the $\mathbf{k} = 0$ state are the condensate, which is the ground state of the system. Quasiparticles with $\mathbf{k} \neq 0$ are excitations, and H_0 is the effective Hamiltonian of these excitations. It contains combinations of operators such as $C_{\mathbf{k}}^\dagger C_{-\mathbf{k}}^\dagger$ which correspond to the excitation of two particles from the condensate. Since momentum is conserved, they must have $\mathbf{k}$ and $-\mathbf{k}$. Similarly, the term with $C_{\mathbf{k}} C_{-\mathbf{k}}$ is the destruction of two quasiparticles with $\mathbf{k}$ and $-\mathbf{k}$ when both are returned to the condensate. The terms with $C_{\mathbf{k}}^\dagger C_{\mathbf{k}}$ correspond to the scattering of quasiparticles by the condensate, and there are both direct and exchange processes which lead to the terms with $V(0)$ and $V(\mathbf{k})$, respectively.

The Hamiltonian (11.25) for H_0 is solved to obtain the eigenvalues and eigenstates of the system. This solution is approximate, since it omits the effects of the potential V. The approximation of neglecting V and using only the eigenstates of H_0 is the pairing theory of Bogoliubov. The Hamiltonian is rather easy to diagonalize. Each wave vector state is treated independently, and each of these is just the problem encountered in Problem 3 in Chapter 1. Introduce a set of harmonic oscillator coordinates $Q_{\mathbf{k}}$ and $P_{\mathbf{k}}$, and the Hamiltonian is an example where the interactions merely shift the eigenfrequency of the oscillator:

$$Q_{\mathbf{k}} = \frac{1}{\sqrt{2\varepsilon_{\mathbf{k}}}} (C_{\mathbf{k}} + C_{-\mathbf{k}}^\dagger) \tag{11.26}$$

$$P_{\mathbf{k}} = i\sqrt{\frac{\varepsilon_{\mathbf{k}}}{2}} (C_{\mathbf{k}}^\dagger - C_{-\mathbf{k}}) \tag{11.27}$$

$$H_0 = \tfrac{1}{2} \sum_{\mathbf{k}}' \{P_{\mathbf{k}} P_{\mathbf{k}} + Q_{\mathbf{k}} Q_{\mathbf{k}} [\varepsilon_{\mathbf{k}}^2 + 2\rho_0 \varepsilon_{\mathbf{k}} V(\mathbf{k})]\} \tag{11.28}$$

$$H_0 = \tfrac{1}{2} \sum_{\mathbf{k}}' \{P_{\mathbf{k}} P_{\mathbf{k}} + \Omega_{\mathbf{k}}^2 Q_{\mathbf{k}} Q_{\mathbf{k}}\}$$

$$\Omega_{\mathbf{k}}^2 = \varepsilon_{\mathbf{k}} [\varepsilon_{\mathbf{k}} + 2\rho_0 V(\mathbf{k})] \tag{11.29}$$

The first two terms in H_0 come from $C_{\mathbf{k}}^\dagger C_{\mathbf{k}}$, while the last term comes from the second term in the Hamiltonian H_0. The terms multiplying $Q_{\mathbf{k}} Q_{-\mathbf{k}}$ are combined into the square of an effective frequency $\Omega_{\mathbf{k}}$. The bosons behave as a harmonic oscillator system, for each wave

vector $\mathbf{k}$, at the new frequency $\Omega_{\mathbf{k}}$. The Hamiltonian is diagonalized by defining a new set of creation and destruction operators

$$H_0 = \sum_{\mathbf{k}}' \Omega_{\mathbf{k}} \alpha_{\mathbf{k}}^\dagger \alpha_{\mathbf{k}} \tag{11.30}$$

$$Q_{\mathbf{k}} = \frac{1}{\sqrt{2\Omega_{\mathbf{k}}}} (\alpha_{\mathbf{k}} + \alpha_{-\mathbf{k}}^\dagger) \tag{11.31}$$

$$P_{\mathbf{k}} = i\sqrt{\frac{\Omega_{\mathbf{k}}}{2}} (\alpha_{\mathbf{k}}^\dagger - \alpha_{-\mathbf{k}}) \tag{11.32}$$

In the pairing theory, the energy $\Omega_{\mathbf{k}}$ is that of the quasiparticles, and $\alpha_{\mathbf{k}}$ are the operators which describe the destruction of these quasiparticles.

The equation for the quasiparticle energy $\Omega_{\mathbf{k}} = \{\varepsilon_{\mathbf{k}}[\varepsilon_{\mathbf{k}} + 2\rho_0 V(\mathbf{k})]\}^{1/2}$ makes no sense whatever, since the Fourier transform $V(\mathbf{k})$ does not exist. It should be replaced by a T matrix or reaction matrix, and the latter is correct. According to Eq. (4.31), for each angular momentum state l one can replace the potential by the on-shell reaction matrix

$$V(\mathbf{k}) \to 4\pi R_l(k, k) = -\frac{2\pi \tan[\delta_l(k)]}{\mu k} \tag{11.33}$$

where $\mu = m/2$ is the reduced mass for the scattering of two identical particles of mass m. The reaction matrix has the important feature that it goes to a constant in the limit where $k \to 0$. The phase shifts behave in this limit as

$$\lim_{k\to 0} \delta_l(k) = m_l \pi - a k^{2l+1} \tag{11.34}$$

$$\lim_{k\to 0} \frac{\tan(\delta_l)}{k} = -a k^{2l} \tag{11.35}$$

where m_l is an integer which denotes the number of bound states in that scattering channel. The quantity a is called a scattering length.

The largest term is from the s-wave channel, in which $R_0(0, 0) = 2\pi a/\mu$, while $R_l(0, 0) = 0$ for $l > 0$. Only the s-wave needs to be included in the dispersion relation of the excitations as $k \to 0$

$$V(\mathbf{k}) \to \frac{2\pi a}{\mu} = \frac{4\pi a}{m} \tag{11.36}$$

$$\Omega_{\mathbf{k}} \to \sqrt{2\varepsilon_{\mathbf{k}} \rho_0 (4\pi a/m)} = ck \tag{11.37}$$

$$c = \frac{\hbar}{m}\sqrt{4\pi a \rho_0} \tag{11.38}$$

The scattering length a is positive for repulsive potentials. For a hard sphere potential of radius σ, then $\sigma = a$.

In the earlier days of helium theory, the helium-helium interaction was approximated as a hard core at the radius $a = \sigma = 0.265\,\text{nm}$ of the Lennard-Jones potential. All the other

parameters are known since ρ_0 is the ^{4}He density. The velocity of the excitation may be calculated:

$$a = 0.265\,\text{nm} \tag{11.39}$$

$$m = 6.648 \times 10^{-27}\,\text{kg} \tag{11.40}$$

$$\rho_0 = 2.2 \times 10^{22}\,\text{atoms/cm}^3 \tag{11.41}$$

$$c = 130\,\text{m/s} \tag{11.42}$$

The predicted velocity is close to the speed of ordinary sound in helium, which is $c_s = 220\,\text{m/s}$. This agreement was viewed as a great success of the pairing theory, since the collective excitations of the liquid at long wavelength must certainly be the longitudinal sound waves.

Unfortunately, this agreement is entirely superfluous. The interaction potential between two ^{4}He particles has attractive regions, and the phase shifts for this potential show there is a low-energy scattering resonance. The exact value of the scattering length a is uncertain because it is very sensitive to the potential. The expected values are more in the range of 50σ rather than σ. A value of $a \sim 50\sigma$ will make the calculated speed of sound be 4 times larger than the experiment. Of course, the sign of a could be negative, in which case the predicted frequencies are complex, and the vibrations are unstable. A possible way out of this difficulty is to do the theory more carefully. This was done by Brown and Coopersmith (1969), who found even worse agreement with experiment. Their calculation represented all direct and exchange scattering by reaction matrices, and they self-consistently found the condensate fraction f_0 and the chemical potential. They obtained a speed of sound about 10 times that of the experiment and also predicted a roton energy about 10 times too big. Indeed, their theoretical excitation spectrum of ^{4}He is similar to the actual one but scaled up by a factor of 10. This calculation was done carefully and correctly. Its lack of success is caused by the failure of the pairing approximation. Indeed, the history of the pairing theory is that each time the theoretical calculations were done better, the answers got worse. The conclusion is that the pairing theory is bad and should be discarded as a way of calculating the excitation spectrum of ^{4}He. Some of the concepts are retained, in particular that of the condensate.

11.1.3. Off-Diagonal Long-Range Order

The modern theory of superfluidity in ^{4}He treats the liquid as a highly correlated and strongly interacting system. The pairing theory is completely abandoned, as is the representation of operators in plane-wave basis states. The one concept carried over is the condensate, or zero-momentum state, in which there is a nonzero fraction of the particles in the superfluid. There is an obvious problem of trying to decide the definition of "zero-momentum state" in a basis set which is not plane waves. It is really part of a larger problem of trying to decide what is meant by superfluidity and Bose–Einstein condensation in a system which is strongly interacting and highly correlated. The whole concept of Bose–Einstein condensation is based on particles occupying zero-momentum states, or not occupying them, and another basis set is needed for a more general definition. The method of doing this was introduced by Penrose (1951), and his ideas are described. Yang (1962) suggested the name of *off-diagonal long-range order* (ODLRO) for the type of ordering introduced by Penrose and also discussed by Penrose and Onsager (1956).

ODLRO was introduced as the type of ordering for superfluids, such as Bose–Einstein condensation in ^{4}He, and electron pairs in superconductors. It is distinguished from diagonal long-range order (DLRO), which is the usual ordering one finds, for example, in crystalline solids. When ^{4}He atoms arrange themselves into a solid, as happens under pressure, they are exhibiting DLRO. But when they go superfluid, they exhibit ODLRO. The distinction between these two processes stems from the different behavior in the density matrix for each type of ordering. These differences will be explained. Further discussion is given by Kohn and Sherrington (1970).

The system has $N \sim 10^{23}$ identical bosons, which are spinless. Their ground state is described by a many-particle wave function $\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$. The actual method of calculating this wave function, or at least some of its properties, is given in the following sections. The present discussion only needs to assume that the wave function exists. The subscript zero on Ψ_0 indicates it is for the ground state. Note that there is no assumption that each atom is in the ground state (i.e., $\mathbf{k} = 0$) but that the system is in the ground state. When the particles are strongly interacting and highly correlated, they will spend little time with $\mathbf{k} = 0$ and will fluctuate between many momentum states. The ground state is the lowest possible energy state for the whole liquid. It is not a static or rigid structure, since each atom fluctuates with zero-point energy. The present estimates for ^{4}He at $T = 1$ K give the average kinetic energy per particle as 15 K and the average potential energy as -22 K, so the average binding energy is 7 K. The large value of average kinetic energy shows the large amount of zero-point motion in the fluid, even at low temperature, which comes from the quantum nature of the fluid. The classical estimate $\langle \text{K.E.} \rangle = 3k_BT/2$ is obviously inaccurate. The ground state wave function $\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ for ^{4}He describes a system with a large amount of zero-point motion.

The square of the wave function gives the probability density for finding particles at positions $\mathbf{r}_j$ in the system and is called the diagonal density matrix:

$$\rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = |\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)|^2 \tag{11.43}$$

The subscript N indicates that it applies to N particles. ρ_N is normalized so that the integral over all coordinates gives unity:

$$1 = \int d^3r_1 \cdots d^3r_N \rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{11.44}$$

The one-particle density matrix is obtained from ρ_N by integrating over all but one coordinate:

$$\rho_1(\mathbf{r}_1) = \int d^3r_2 d^3r_3 \cdots d^3r_N \rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{11.45}$$

$$1 = \int d^3r_1 \rho_1(\mathbf{r}_1) \tag{11.46}$$

The notation is a bit confusing, since $\rho_1(\mathbf{r}_1)$ is not the probability in a one-particle system but rather the probability of one particle being at $\mathbf{r}_1$ in the N-particle system. Since all points are equivalent, $\rho_1(\mathbf{r}_1)$ is really independent of position—neglecting edges—so $\rho_1 = 1/v$, where v is the volume of the system.

Another useful quantity is the two-particle density matrix $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ which is obtained from ρ_N by integrating over all but two coordinates:

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \int d^3r_3 d^3r_4 \cdots d^3r_N \rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{11.47}$$

$$\frac{1}{v} = \int d^3r_2 \rho_2(\mathbf{r}_1, \mathbf{r}_2) \tag{11.48}$$

Now $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ is the probability that any boson is at $\mathbf{r}_2$ if there is one at $\mathbf{r}_1$. In a homogeneous system it must depend only on the difference $\mathbf{r}_1 - \mathbf{r}_2$. In fact, it must be proportional to the pair distribution function $g(r)$, and this relationship is

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{v^2} g(\mathbf{r}_1 - \mathbf{r}_2) \tag{11.49}$$

which follows from the fact that $g(r)$ is normalized to be unity at large r.

Crystalline solids have DLRO, which is indicated by structure in $g(\mathbf{r})$. Since the atoms are regularly spaced, $g(\mathbf{r})$ has large values where the atoms are located and is zero otherwise. Thermal vibrations smear out the answer.

Penrose (1951) used the idea of a general density matrix which is defined as the product of two wave functions with different coordinates:

$$\tilde{\rho}_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \mathbf{r}_1', \mathbf{r}_2', \ldots, \mathbf{r}_N') \tag{11.50}$$

$$= \Psi_0^*(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)\Psi_0(\mathbf{r}_1', \mathbf{r}_2', \ldots, \mathbf{r}_N') \tag{11.51}$$

This quantity is denoted by $\tilde{\rho}$, where the tilde is to distinguish it from the diagonal density matrix introduced earlier. Of course, they are identical if the two sets of coordinates in $\tilde{\rho}$ are set equal: $\tilde{\rho}_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$. The concept of ODLRO is contained in the function $\tilde{\rho}(\mathbf{r}_1, \mathbf{r}_1')$ obtained from $\tilde{\rho}_N$ when all but one set of coordinates are set equal and averaged over

$$\tilde{\rho}_1(\mathbf{r}_1, \mathbf{r}_1') = \int d^3r_2 d^3r_3 d^3r_4 \cdots d^3r_N \tilde{\rho}_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \mathbf{r}_1', \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

$$\rho_1(\mathbf{r}_1, \mathbf{r}_1) = \rho_1(\mathbf{r}_1) = \frac{1}{v} \tag{11.52}$$

where $\tilde{\rho}_1(\mathbf{r}_1, \mathbf{r}_1')$ becomes the diagonal density matrix ρ_1 when $\mathbf{r}_1 = \mathbf{r}_2$. In a liquid, the dependence on $\mathbf{r}_1$ and $\mathbf{r}_2$ can only be their difference, since there is no absolute frame of reference. It is convenient to define the quantity

$$R(\mathbf{r}_1 - \mathbf{r}_1') = v\tilde{\rho}_1(\mathbf{r}_1, \mathbf{r}_1') \tag{11.53}$$

$$R(0) = 1 \tag{11.54}$$

which is normalized to unity at $\mathbf{r} = 0$. $R(r)$ is the function which is important in understanding ODLRO.

Some insight into $\tilde{\rho}(\mathbf{r}_1, \mathbf{r}_1')$ is obtained by reconsidering the techniques used in the weakly interacting systems. There a one-particle state function is defined in the plane-wave representation as

$$\Phi(\mathbf{r}) = \frac{1}{\sqrt{v}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} C_{\mathbf{k}} \tag{11.55}$$

$$C_{\mathbf{k}} = \frac{1}{\sqrt{v}} \int d^3r e^{-i\mathbf{k}\cdot\mathbf{r}} \Phi(\mathbf{r}) \tag{11.56}$$

The number of particles in state $\mathbf{k}$ is

$$n_{\mathbf{k}} = \langle C_{\mathbf{k}}^{\dagger} C_{\mathbf{k}} \rangle = \frac{1}{v} \int d^3r d^3r' e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \langle \Phi^{\dagger}(\mathbf{r})\Phi(\mathbf{r}') \rangle \tag{11.57}$$

$n_{\mathbf{k}}$ is found to be the Fourier transform of the quantity $\langle \Phi^{\dagger}(\mathbf{r})\Phi(\mathbf{r}') \rangle$, which must be a function of $\mathbf{r} - \mathbf{r}'$. Recall what this average means. Take a particle and find its wave function at two different points $\mathbf{r}$ and $\mathbf{r}'$. This product is averaged, which must be taken over the other particles and their positions. This procedure is exactly the one which was used to obtain $\tilde{\rho}_1(\mathbf{r}_1, \mathbf{r}_1')$, except now it is recognized from the outset that the particle is part of a many-particle system. The procedure is to take a product wave function with one particle at two points $\mathbf{r}_1$ and $\mathbf{r}_1'$ and average it over the other particles and their positions. The quantity $\tilde{\rho}_1(\mathbf{r}_1, \mathbf{r}_1')$ or its equivalent $R(\mathbf{r}_1 - \mathbf{r}_2)$, is the many-body definition of the quantity $\langle \Phi^{\dagger}(\mathbf{r})\Phi(\mathbf{r}') \rangle$. It is also important to appreciate that $R(r)$ is quite a different function from the pair distribution function $g(r)$ and the two are not related. The quantity $n_{\mathbf{k}}$ is also related to the Wigner distribution function where $n_{\mathbf{k}}(R, T) = \int d\omega f(\mathbf{k}, \omega; R, T)$.

The quantity $n_{\mathbf{k}}$ is the number of particles, on the average, in the momentum state $\mathbf{k}$. It is an important quantity in Bose–Einstein condensation, since one expects that one momentum state $\mathbf{k}_0$ will have a macroscopic occupation such that $n_{\mathbf{k}_0} = f_0 N$, where f_0 is still the fraction of particles in the condensate. This fraction is of order unity, rather than $O(1/N)$. Normally $\mathbf{k}_0 = 0$, which applies when the fluid is at rest. There are circumstances when another state has the macroscopic occupation, e.g., when the fluid is flowing at a uniform rate or when it is rotating. For $\mathbf{k} \neq \mathbf{k}_0$,

$$n_{\mathbf{k}} = \int \frac{d\varepsilon}{2\pi} n_B(\varepsilon) A(\mathbf{k}, \varepsilon) \tag{11.58}$$

where $A(\mathbf{k}, \varepsilon)$ is the spectral function for the boson particles.

The procedure for finding $n_{\mathbf{k}}$ is the same as in the free-particle case (11.57). There one took the Fourier transform of $\langle \Phi^{\dagger}(\mathbf{r})\Phi(\mathbf{r}') \rangle$, while now one takes the Fourier transform of the equivalent quantity $R(\mathbf{r} - \mathbf{r}')$. One difficulty with this procedure is that $R(r)$ may not possess a Fourier transform. For example, how does it behave as $r \to \infty$? For fluids at rest, so that the macroscopic occupation is in the zero-momentum state, $R(r)$ goes to a constant as $\mathbf{r} \to \infty$. Call this constant $R(\infty)$. The Fourier transform is

$$n_{\mathbf{k}} = \rho_0 \int d^3r R(r) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{11.59}$$

$$= \rho_0 R(\infty)(2\pi)^3 \delta^3(\mathbf{r}) + \rho_0 \int d^3r [R(r) - R(\infty)] e^{i\mathbf{k}\cdot\mathbf{r}} \tag{11.60}$$

$$= N_0 \delta_{\mathbf{k}=0} + \rho_0 \int d^3r [R(r) - R(\infty)] e^{i\mathbf{k}\cdot\mathbf{r}} \tag{11.61}$$

The density is $\rho_0 = N/v$, $N_0 = f_0 N$. There is a delta function term at $\mathbf{k} = 0$ whose amplitude is $\rho_0 R(\infty)$. The quantity $R(\infty) = f_0$ is the fraction of particles in the zero-momentum state. This fraction is the "order" which exists in the off-diagonal density matrix $R(r)$. For $\mathbf{k} \neq 0$, the momentum distribution of particles $n_{\mathbf{k}}$ is found from the Fourier transform of $[R(r) - f_0]$.

Figure 11.3 shows the function $R(r)$ calculated for liquid ^{4}He by McMillan (1965) using Monte Carlo techniques. He used the Lennard-Jones potential between helium atoms and found $f_0 = 0.11$. His fraction is similar to that obtained earlier by Penrose and Onsager, who got $f_0 = 0.08$ for a gas of hard spheres. The latter estimate is also based on a Monte Carlo calculation. It appears, unfortunately, that reliable results can be obtained in this field only by

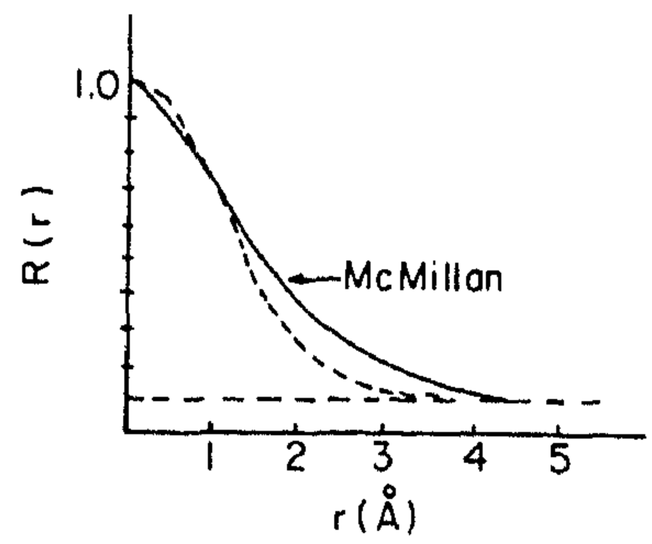


FIGURE 11.3 Single-particle density marix as a function of separation. The solid line is the Monte Carlo calculation of McMillan (1965), which as asymptotically approaches a density fraction of $f_0 = 0.11$. The dashed curve is the Gaussian approximation.

extensive computer calculations. There have been many calculations of these quantities; e.g., see Francis *et al.* (1970). Figure 11.3 shows that $R(r)$ starts at unity and falls smoothly to its asymptotic value, which it reaches at about $r = 0.4$ nm. This distance is rather short, since it is only 1.5 atomic diameters.

The quantity $f_0 = R(\infty)$ is the fraction of time a particle spends in the condensate. Alternately, it is also the fraction of particles in this state at any one time. However, in calculating this quantity, it has not been assumed that the system is superfluid. The value of $R(\infty)$ is found by solving the dynamical properties of normal fluids, which may seem paradoxical and is probably best understood in the next section which gives more details of the method. Normal fluids do not have a condensate and have $f_0 = R(\infty) = 0$. However, the point is that most methods of calculating f_0 are based on the properties of normal fluids. For example, the estimate of Penrose and Onsager came from the equation of state of a classical hard sphere gas, which is certainly not a superfluid system.

11.1.4. Correlated Basis Functions

Correlated basis functions (CBFs) are the type of wave function most often employed in the study of the ground state properties of ^{4}He. They have the form

$$\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = L_N \exp\left[-\sum_{i>j} u(\mathbf{r}_i - \mathbf{r}_j)\right] \tag{11.62}$$

where L_N is a normalization constant which is specified by satisfying the integral (11.44). They were used by Bijl (1940), Dingle (1949), and Jastrow (1955) and are sometimes named after these authors; different writers preferring various combinations of their names. The CBFs have several advantages which make them desirable for a description of ^{4}He. The most important is that the wave function possesses the necessary symmetry of being symmetric under the exchange of any two coordinates. This symmetry is automatically satisfied by the summation of pairwise correlations. Another advantage is that Ψ_0 has a simple form for which one needs to determine only one function $u(r)$, which is usually found variationally by minimizing the ground state energy.

A third advantage of CBFs is that the diagonal density matrix has a mathematical form which is identical to another problem which has been studied extensively—the classical fluid. For a classical fluid, the kinetic energy terms are irrelevant for determining particle

correlations, and only the potential energy $V(\mathbf{r}_i - \mathbf{r}_j)$ between particles i and j is important. The density matrices for CBFs and the classical fluid are

$$\text{CBF:} \quad \rho N = |\Psi_0|^2 = L_N^2 \exp\left[-2\sum_{i>j} u(\mathbf{r}_i - \mathbf{r}_j)\right] \tag{11.63}$$

$$\text{Classical:} \quad \rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = Q_n \exp\left[-\beta\sum_{i>j} V(\mathbf{r}_i - \mathbf{r}_j)\right] \tag{11.64}$$

These have the same mathematical form when setting $2u(\mathbf{r}) = \beta V(\mathbf{r})$. The classical problem has been extensively studied, and accurate calculational techniques have been developed which work well for liquids of neutral atoms such as the rare gases. The most successful methods are based on the Percus–Yevick equation (see Percus, 1964). The availability of this successful computational technology is one reason for the popularity and success of CBFs.

One important quantity is the off-diagonal density matrix $R(r)$, which is defined in terms of CBFs as

$$R(\mathbf{r}_1 - \mathbf{r}_1') = vL_N^2 \int d^3r_2 d^3r_3 \cdots d^3r_N \times \exp\left[-\sum_{j=2}^{N} [u(\mathbf{r}_1 - \mathbf{r}_j) + u(\mathbf{r}_1' - \mathbf{r}_j)] - 2\sum_{i>j=2} u(\mathbf{r}_i - \mathbf{r}_j)\right] \tag{11.65}$$

Let us determine the classical analog of this expression. That is, regard $u(\mathbf{r}_i - \mathbf{r}_j)$ as the effective "potential" $\beta V(\mathbf{r}_i - \mathbf{r}_j)/2$ between particles and see where that leads us. Obviously there is a system of $N-1$ particles which mutually interact with each other with the "potential" $2u(r)$. There is a particle at $\mathbf{r}_1$ which interacts with these $N-1$ particles with the different potential $u(r)$. Another particle at $\mathbf{r}_1'$ also interacts with the $N-1$ particles with the same relative potential $u(r)$, so that it is identical to the particle at $\mathbf{r}_1$. Finally, the particles at $\mathbf{r}_1$ and $\mathbf{r}_1'$ have no mutual interaction. This description is of a system of two impurity particles, at $\mathbf{r}_1$ and $\mathbf{r}_1'$, in a system of $N-1$ other particles. The impurities are different, since they have a different interaction with the $N-1$ particles and no interaction between themselves. This classical system can be solved by using the Percus-Yevick equations for binary mixtures, where one constituent is very dilute. This technique was used by Francis *et al.*, (1970) for their investigation of superfluid ^{4}He. It is another example of where classical equations are solved to determine the properties of the superfluid.

The ground state properties which are calculated with CBFs include the condensate fraction f_0, the momentum distribution $n_{\mathbf{k}}, R(r), g(r)$, and $S(k)$. Another property is the ground state energy per atom, which is usually presented as the separate contributions from potential energy (P.E.) and kinetic energy (K.E.). The potential energy per atom is given by

$$\frac{\langle \text{P.E.} \rangle}{N} = \frac{1}{N}\sum_{i>j} \int d^3r_1 \cdots d^3r_N V(\mathbf{r}_i - \mathbf{r}_j)\rho_N(\mathbf{r}_1 \cdots \mathbf{r}_N) \tag{11.66}$$

The density matrix $\rho_N(\mathbf{r}_1 \cdots \mathbf{r}_N)$ is unchanged when any pair of coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$ are interchanged. With this fact, one can show that the average value of each potential energy term $V(\mathbf{r}_i - \mathbf{r}_j)$ is the same as any other term $V(\mathbf{r}_i' - \mathbf{r}_j')$. The average potential of any particle

is the same as any other. The summation over i and j gives $N(N-1)/2$, which is the number of pairs, and all pairs contribute equally:

$$\frac{\langle \text{P.E.}\rangle}{N} = \tfrac{1}{2}(N-1)\int d^3r_1 d^3r_2 V(\mathbf{r}_1-\mathbf{r}_2)\int d^3r_3 d^3r_4\cdots d^3r_N \rho_N(\mathbf{r}_1\cdots\mathbf{r}_N)$$

The integration over all variables except $\mathbf{r}_1$ and $\mathbf{r}_2$ produces the reduced density matrix $\rho_2(\mathbf{r}_1,\mathbf{r}_2) = g(\mathbf{r}_1-\mathbf{r}_2)/v^2$. Then changing to relative coordinates $\mathbf{r} = (\mathbf{r}_1-\mathbf{r}_2)$, $\mathbf{R} = (\mathbf{r}_1+\mathbf{r}_2)/2$, there is no dependence on $\mathbf{R}$, so its integral yields v. The final answer is

$$\frac{\langle \text{P.E.}\rangle}{N} = \frac{\rho_0}{2}\int d^3r V(r) g(r) \tag{11.67}$$

The average potential energy is obtained from the pair distribution function. This result is obvious for a system with pairwise interaction between particles. It does not depend on CBFs but applies to any liquid and is an exact identity. The CBFs are used to calculate $g(r)$, which is then used in the evaluation of the average potential energy per particle. The factor of $\frac{1}{2}$ occurs because each pair interaction is shared between two atoms.

The average kinetic energy per particle is obtained from the fundamental definition

$$\frac{\langle \text{K.E.}\rangle}{N} = \frac{\hbar^2}{2mN}\sum_j \int d^3r_1\cdots d^3r_N |\nabla_j \Psi_0(\mathbf{r}_1,\ldots,\mathbf{r}_N)|^2 \tag{11.68}$$

The result for each term j in the summation is identical, so the sum gives N times the result for one term:

$$\frac{\langle \text{K.E.}\rangle}{N} = \frac{\hbar^2}{2m}\int d^3r_1\cdots d^3r_N |\nabla_1 \Psi_0(\mathbf{r}_1,\ldots,\mathbf{r}_N)|^2 \tag{11.69}$$

The evaluation of this quantity depends sensitively on the use of CBFs. The gradient of the wave function produces the gradient acting upon $u(r)$:

$$\nabla_1\Psi_0(\mathbf{r}_1,\ldots,\mathbf{r}_N) = L_N\nabla_1 \exp\left[-\sum_{i>j} u(\mathbf{r}_i-\mathbf{r}_j)\right] = -\Psi_0\sum_{j=2}^{N}\nabla_1 u(\mathbf{r}_1-\mathbf{r}_j)$$

This expression is squared in the evaluation of the average kinetic energy. At first the evaluation appears to involve three-particle correlations. One can integrate all but the three coordinates r_1, r_j, r_m, which produces the reduced density matrix $\rho_3(\mathbf{r}_1,\mathbf{r}_j,\mathbf{r}_m)$. It would be difficult to evaluate ρ_3 accurately, since the knowledge of three-particle correlations is imperfect. However, McMillan (1965) suggested a method of avoiding this problem and expressing the result in terms of only two-particle correlations. His method is to write one derivative as acting upon the density matrix itself:

$$\begin{aligned}\rho_N\sum_{j=2}^{N}\nabla_1 u(\mathbf{r}_1-\mathbf{r}_j) &= -\tfrac{1}{2}\nabla_1\rho_N(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)\\ \frac{\langle K.E.\rangle}{N} &= -\frac{\hbar^2}{4m}\sum_j\int d^3r_1 d^3r_j \nabla_1 u(\mathbf{r}_1-\mathbf{r}_j)\cdot\nabla_1\rho_2(\mathbf{r}_1,\mathbf{r}_j)\end{aligned} \tag{11.70}$$

All coordinates were integrated over except $\mathbf{r}_1$ and $\mathbf{r}_j$. Eliminate the summation over j since each term contributes identically:

$$\frac{\langle K.E.\rangle}{N} = -(N-1)\frac{\hbar^2}{4m}\int d^3r_1 d^3r_2 \nabla_1 u(\mathbf{r}_1 - \mathbf{r}_2)\cdot\nabla_1 \rho_2(\mathbf{r}_1, \mathbf{r}_2) \tag{11.71}$$

Again changing integration variables to relative coordinates, the final form of the result is

$$\frac{\langle K.E.\rangle}{N} = -\rho_0 \frac{\hbar^2}{4m}\int d^3r \frac{du(r)}{dr}\frac{dg(r)}{dr} \tag{11.72}$$

The average kinetic energy is expressed as an integral over the derivative of $u(r)$ and $g(r)$. Neither quantity has an angular dependence, so the gradient acts only upon the radial variable. This form is used for the evaluation of the average kinetic energy per particle. The numerical results obtained by different investigators vary somewhat, depending on the details. Generally the results show that the average potential energy per particle has the magnitude of 20–22 K, while the average kinetic energy per particle is 14–15 K, and the net binding energy per particle is 6–7 K. These values depend on the particle density.

So far the form of $u(r)$ in CBFs has not been specified. That will be done now. Usually the contribution of $u(r)$ is divided into a short-range component and a long-range component. The short-range component is the most straightforward and is discussed first. One expects that the atoms do not interpenetrate. There must be a term in $u(r)$ which keeps them apart, which is done by having a term in $u(r)$ which becomes very large as $r \to 0$, which makes $\exp(-u) \to 0$. The form of this term can be deduced from Schrödinger's equation and the repulsive part of the interparticle potential. Take a two-particle system, so that $u(r)$ describes only the effects of correlation between the motion of two particles. In center-of-mass coordinates, the relative motion of the two helium atoms is described by the following CBF wave functions:

$$\left\{-\frac{\hbar^2}{2\mu}\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{l(l+1)}{r^2}\right] + V(r)\right\}e^{-u(r)} = Ee^{-u(r)} \tag{11.73}$$

$$e^{-u(r)}\left\{-\frac{\hbar^2}{2\mu}\left[-\frac{d^2u}{dr^2} + \left(\frac{du}{dr}\right)^2 - \frac{2}{r}\frac{du}{dr} - \frac{l(l+1)}{r^2}\right] + V(r) - E\right\} = 0 \tag{11.74}$$

If $u(r)$ is diverging sharply as $r \to 0$, the dominant term among all the derivatives is $(du/dr)^2$. For example, if $u \propto r^{-n}$, then $(du/dr)^2 \propto r^{-2n-2}$, while the second derivative diverges only as r^{-n-2}. The two most divergent terms must cancel, where one is from the kinetic energy and the other from the potential energy:

$$\frac{du}{dr} = -\left[\frac{2\mu}{\hbar^2}V(r)\right]^{1/2} \tag{11.75}$$

Students of physics should recognize this result and derivation. It is just the Wentzel–Kramers–Brillouin–Jeffreys (WKBJ) wave function for a particle penetrating into a repulsive

potential: as a specific example, in the Lennard-Jones potential the repulsive part is $4\varepsilon(\sigma/r)^{12}$. In this case the short-range part of the CBF must be

$$\frac{du}{dr} = -\left(\frac{8\mu\varepsilon}{\hbar^2}\right)^{1/2}\left(\frac{\sigma}{r}\right)^6 \tag{11.76}$$

$$u(r) = \left(\frac{8\mu\varepsilon}{\hbar^2}\right)^{1/2}\frac{\sigma^6}{5r^5} \equiv \left(\frac{a}{r}\right)^5 \tag{11.77}$$

$$a = \sigma\left(\frac{8\mu\varepsilon\sigma^2}{25\hbar^2}\right)^{1/10} \tag{11.78}$$

For the potential between two ^{4}He atoms, $\varepsilon = 1.484 \times 10^{-22}$ J, $\sigma = 0.22648$ nm, and the numerical result is $a = 1.01\sigma$. The constant a is nearly identical to the hard sphere diameter σ. McMillan (1965) used the Lennard-Jones potential for his variational calculation and included no long-range potential. He tried correlations of the form $u(r) = (a/r)^n$, where a and n were both variational parameters. He used $\sigma = 0.2556$ nm so $a = 1.01\sigma = 0.259$ nm is our prediction. His minimum energy was found with $n = 5$ and $a = 0.261$ nm. These values are nearly identical to the results using the WKBJ between two particles. The short-range part of the CBF can be obtained from the wave function for a particle penetrating into a repulsive potential. This part of $u(r)$ is determined once the potential is specified.

The long-range part of $u(r)$ is more subtle. There is a term which falls off as $u(r) \propto r^{-2}$ as $r \to \infty$. This contribution arises from the zero-point motion of the phonons. Bogoliubov and Zubarev (1955) showed, in the pairing theory, that the phonons lead to long-range correlations between the motion of particles in the liquid. Reatto and Chester (1967) showed that these fluctuations lead to long-range contributions to $u(r)$. They recommended $u(r) = u_s(r) + u_l(r)$, where the short-range result $u_s(r) = (a/r)^5$, and the long-range part is

$$u_l(r) = \frac{mc}{2\pi^2\rho_0}\frac{1}{r^2 + r_c^2} \tag{11.79}$$

where the cut-off is $r_c \approx 0.20$ nm. The parameter c is the speed of sound. The function $u_l(r)$ goes to a constant as $r \to 0$ and only becomes proportional to r^{-2} at large distances. An inverse square law is a rather slow fall off with distance and is the long-range behavior.

The same $u(r)$ was used by Francis *et al.* (1970) in the numerical study referred to previously. The inclusion of the long-range correlations will change $R(r)$ at large r and $n_{\mathbf{k}}$ at small k. In particular, the distribution $n_{\mathbf{k}}$ diverges as $\mathbf{k} \to 0$ as a result of the long-range correlations. The divergence is shown in Fig. 11.4 where the results for $n_{\mathbf{k}}$ of Francis *et al.* are shown for both the short-range correlations only (SR) and the short plus long-range correlations (LR). Their short-range result is nearly identical to that of McMillan. This agreement is expected, since they used the same interparticle potential but a different method of obtaining results, i.e., the Percus-Yevick equations for mixtures. The long-range correlations have a dramatic effect upon the theoretical values of $n_{\mathbf{k}}$. This low k divergence is found in the experiments, as discussed by Griffin (1985), and by Svensson and Sears (1986).

Amid all these equations, it is worthwhile to gain a physical picture of the particle motions on the microscopic level. The average potential energy of 20 to 22 K per atom can be achieved by having four neighbors at the minimum well depth of 10 K, so the average potential energy per particle is 20 K. However, an inspection of $g(r)$ shows that it peaks at larger values of r, and the neighboring shell of atoms is farther away than the minimum well depth at 0.3 nm. The potential energy comes from having about eight neighbors, each with an

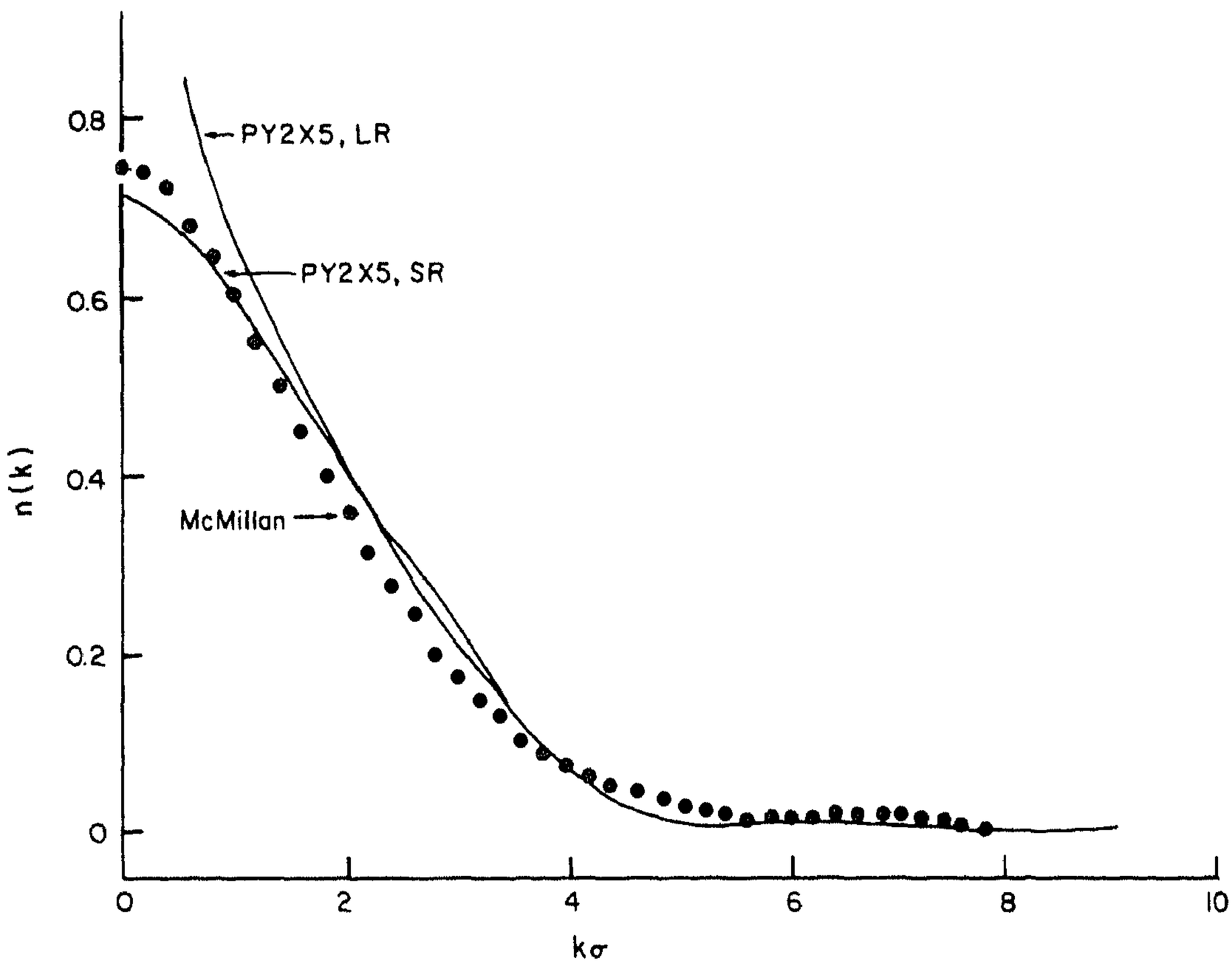


FIGURE 11.4 Theoretical calculations of the momentum distribution of atoms in liquid ^{4}He. The points are due to McMillan (1965), and the solid line is from Francis *et al.*, who used the Percus-Yevick equation both with and without long-range order. *Source:* Francis *et al.*, (1970) (used with permission).

average interaction energy of 5 K, which gives the per particle average of 20 K. In a classical fluid such as argon, $g(r)$ peaks at the maximum well depth, so the first neighbor atoms are sitting at the distance of the maximum potential energy. Helium does not behave this way, because the quantum nature of the motion prevents that much localization of the particles, so the first neighbors are farther away on the average.

The average kinetic energy is the more interesting number. The estimates of this quantity range from 10 to 15 K. What kinds of particle motion give this large number? There are two microscopic pictures which come to mind. The first has the particle motion as part of the zero-point motion of the long-wavelength phonons. Here the particles keep their nearest neighbor positions rather fixed, but the system executes long-range fluctuations. This Jell-O model has the system wiggling like an elastic medium. The amount of energy in this zero-point motion can be estimated as

$$(\text{Z.P.E.}) = \frac{\hbar}{2\rho_0}\int \frac{d^3q}{(2\pi)^3}(cq)\Theta(q_c - q) \tag{11.80}$$

$$\approx \frac{\hbar c q_c^4}{16\pi^2\rho_0} \tag{11.81}$$

Here c is the speed of sound and ρ_0 is the particle density. The cutoff wave vector q_c is the maximum value of the excitation spectrum, which is considered to be a sound wave, i.e., over which $\omega_q = cq$ is an accurate approximation. Francis *et al.*, estimate it to be $q_c = 0.5$ Å^{-1}. With this value and the other known parameters for ^{4}He, one can estimate that (Z.P.E.) = 0.33 K. The result 0.3 K is a negligible part of the total kinetic energy per particle. This

low estimate is confirmed by the calculations of Francis *et al.*, who found that the long-range correlations (and fluctuations) changed the average energy per particle only by 0.5 K.

The average kinetic energy per particle does not come from the long-range fluctuations. Instead, it seems to come from the short-range motion. The first peak in $g(r)$ is at 4 Å, so the atom sits in its own space in the liquid, which is roughly a sphere with a radius of about 2 Å. A particle in a spherical box of radius a has a kinetic energy of $\hbar^2\pi^2/2ma^2$. Using $a = 2$ Å and the mass of helium yields a kinetic energy of 15 K, which is just the right magnitude. The kinetic energy comes from the short-range fluctuations of the particle bouncing around inside its own small space in the liquid. In the ground state of the liquid, at very low temperatures, the individual particles are moving rapidly with this motion.

It is also useful to have a simple approximation for $n_{\mathbf{k}}$ and its Fourier transform $R(r)$. Puff and Tenn (1970) observed that neutron scattering experiments, which will be described later, showed that the particle distribution was Gaussian. They suggested the form

$$n_{\mathbf{k}} = N_0\delta_{\mathbf{k}=0} + (1-f_0)\rho_0\left(\frac{2\pi\hbar^2}{mw}\right)^{1/2}\exp(-\varepsilon_{\mathbf{k}}/w) \tag{11.82}$$

One advantage of the Gaussian distribution is that one can calculate quantities easily. For example, the average kinetic energy per particle is simply

$$\frac{\langle \text{K.E.}\rangle}{N} = (1-f_0)\left(\frac{2\pi\hbar^2}{mw}\right)^{1/2}\int\frac{d^3k}{(2\pi)^3}\,\varepsilon_{\mathbf{k}}e^{-\varepsilon_{\mathbf{k}}/w} = \tfrac{3}{2}(1-f_0)w$$

From their fits to the neutron scattering data, they deduced that this average kinetic energy is about 15 K. It is their result we have mentioned several times previously in this chapter. This form has a certain appeal. The classical distribution is also Gaussian, except then $w = k_BT$, and the average kinetic energy is $(3/2)k_BT$. In ^{4}He there is still a Gaussian distribution, but the width w is much larger than predicted by the temperature; one estimates that $w = 11$ K in the limit where $T \to 0$.

The off-diagonal density matrix $R(r)$ is the Fourier transform of $n_{\mathbf{k}}$. When $n_{\mathbf{k}}$ is Gaussian, so is $R(r)$:

$$R(r) = \frac{1}{\rho_0}\int\frac{d^3k}{(2\pi)^3}\,n_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} = f_0 + (1-f_0)\exp(-r^2/b^2) \tag{11.83}$$

$$b^2 = \frac{2\hbar^2}{mw} \tag{11.84}$$

This Gaussian is the dashed curve in Fig. 11.3, compared with the calculations of McMillan. The fit is reasonable, considering that the parameters were taken from the estimate of Puff and Tenn that $w = 11$ K, which gives $b = 1.48$ Å.

It is likely that neither $n_{\mathbf{k}}$ (for $k \neq 0$) nor $R(r)$ are exact Gaussians in the superfluid state and that the fit to this functional form is only an approximation. Nevertheless, it is useful to have approximate analytical forms, and the Gaussian fits best. Puff and Tenn, in fitting the neutron data to a Gaussian plus a delta function at $\mathbf{k} = 0$, estimated that $f_0 = 0.06$ at $T = 1.27$ K. Their estimate is in agreement with later measurements and extrapolates to a value of $f_0 = 10.8\%$ in the limit where $T \to 0$ if one uses the Bose–Einstein extrapolation $f_0(0)[1-(T/T_c)^{3/2}]$.

11.1.5. Experiments on $n_{\mathbf{k}}$

Feynman first suggested that neutron scattering would be an excellent means of investigating the structure of liquid ^{4}He. His remarks were aimed at a measurement of the excitation spectrum. These measurements by neutron scattering have been very successful and are described in the next section. However, in the past three decades another series of neutron scattering experiments have been performed on liquid ^{4}He in order to measure the momentum distribution of particles $n_{\mathbf{k}}$. This latter experiment involves using higher-energy neutrons and analyzing the data in a fashion different from that used to find the excitation energy. The primary objective of this effort is a direct measurement of the condensate fraction f_0.

Hohenberg and Platzman (1966) made the suggestion that inelastic scattering by very energetic neutrons (say 0.1 eV) would provide a measurement of $n_{\mathbf{k}}$ and hence the condensate fraction f_0, which would show up as a delta function on the $n_{\mathbf{k}}$ distribution. Of course, the delta function would be broadened by experimental resolution, but should still be apparent if the estimates of $f_0 = 0.1$ were correct. The reason for using very energetic neutrons is that the sudden approximation becomes valid when the neutron energy is much higher than the kinetic or potential energy of the ^{4}He atom. In the sudden approximation, one has a direct measurement of $n_{\mathbf{k}}$.

In the neutron scattering experiment, the neutrons with initial wave vector $\mathbf{k}_i$ and initial energy $E_i = \hbar^2 k_i^2/2m$ are directed toward the scattering chamber. Some neutrons are scattered and leave the sample with a final wave vector $\mathbf{k}_f = \mathbf{k}_i + \mathbf{Q}$, where $\mathbf{Q}$ is transferred to the liquid. Similarly, the final energy is $E_f = (\mathbf{k}_i - \mathbf{Q})^2/2m = E_i - \omega$, where ω is the energy transferred to the liquid. The scattering cross section for ^{4}He can be expressed in terms of the dynamic liquid structure factor $S(\mathbf{Q}, \omega)$ as

$$\frac{d^2\sigma}{d\Omega d\omega} = \frac{k_f}{k_i}\sigma_0 S(\mathbf{Q}, \omega) \tag{11.85}$$

where σ_0 is a cross section for neutron scattering from a single alpha particle. Equation (11.85) is a general result which is always valid for neutron scattering. Generally a calculation of $S(\mathbf{Q}, \omega)$ is quite complicated. However, in the sudden approximation, it can be approximated by the expression

$$S(\mathbf{Q}, \omega) = \frac{2\pi}{\rho_0}\int \frac{d^3k}{(2\pi)^3} n_{\mathbf{k}}\left[\delta\left(\omega - \frac{Q^2}{2m} - \frac{\mathbf{Q}\cdot\mathbf{k}}{m}\right) - \delta\left(\omega + \frac{Q^2}{2m} + \frac{\mathbf{Q}\cdot\mathbf{k}}{m}\right)\right] \tag{11.86}$$

where $n_{\mathbf{k}}$ is the momentum distribution of particles in the ground state of the liquid. The outline is to first discuss the implications of this result; the derivation is provided later. The momentum distribution $n_{\mathbf{k}}$ is written as the condensate term $N_0\delta_{\mathbf{k}=0}$ plus the other term which is called $\bar{n}_{\mathbf{k}}$. These two are used in (11.86), and the wave vector integrals are done as far as possible ($\varepsilon_Q = Q^2/2m$)

$$n_{\mathbf{k}} = N_0\delta_{\mathbf{k}=0} + \bar{n}_{\mathbf{k}} \tag{11.87}$$

$$S(\mathbf{Q}, \omega) = \bar{S}(\mathbf{Q}, \omega) - \bar{S}(\mathbf{Q}, -\omega) \tag{11.88}$$

$$\begin{aligned} \bar{S}(\mathbf{Q}, \omega) &= 2\pi f_0\delta(\omega - \varepsilon_Q) + \frac{2\pi}{\rho_0}\int \frac{d^3k}{(2\pi)^3}\bar{n}_{\mathbf{k}}\delta\left(\omega - \frac{Q^2}{2m} - \frac{\mathbf{Q}\cdot\mathbf{k}}{m}\right) \\ \bar{S}(\mathbf{Q}, \omega) &= 2\pi f_0\delta(\omega - \varepsilon_Q) + \frac{1}{2\pi\rho_0}\int_0^\infty k^2 dk\bar{n}_{\mathbf{k}}\int_{-1}^{1} dv\delta\left(\omega - \frac{Q^2}{2m} - \frac{Qkv}{m}\right) \end{aligned} \tag{11.89}$$

The angular integral $dv = d(\cos\theta)$ over the delta function equals m/kQ if $k > (m/Q)|\omega - \varepsilon_Q| \equiv k_Q(\omega)$, so the final result is $(\omega > 0)$

$$\bar{S}(\mathbf{Q}, \omega) = 2\pi f_0 \delta(\omega - \varepsilon_Q) + \frac{m}{2\pi Q \rho_0} \int_{k_Q(\omega)}^{\infty} k dk \bar{n}_{\mathbf{k}} \tag{11.90}$$

Equation (11.90) predicts that $S(\mathbf{Q}, \omega)$ for $\omega > 0$ has a delta function contribution at the energy loss $\omega = \varepsilon_Q = Q^2/2m$. The delta function corresponds to the neutron knocking particles out of the condensate, and the fraction of the spectral strength in this process is f_0. Of course, this delta function is broadened by the resolution of the measuring apparatus but should still be observable. The second term in (11.90) gives the scattering from the particles not in the condensate. This term is left as an integral. However, $\bar{n}_{\mathbf{k}}$ may be obtained from data with small numerical scatter by taking the following derivative $(\omega \neq \varepsilon_Q)$:

$$\frac{\partial}{\partial\omega} S(\mathbf{Q}, \omega) = \frac{m^3}{2\pi Q^3} [(\omega - \varepsilon_Q)\bar{n}_{k_Q(\omega)} + (\omega + \varepsilon_Q)\bar{n}_{k_Q(-\omega)}]$$

This equation provides a direct method of measuring $\bar{n}_{\mathbf{k}}$. As pointed out by Puff and Tenn, the Gaussian distribution in (11.82) fits the scattering data well. For this distribution, the integral can be evaluated analytically

$$S(\mathbf{Q}, \omega) = 2\pi f_0 [\delta(\omega - \varepsilon_Q) - \delta(\omega + \varepsilon_Q)] \tag{11.91}$$

$$+ (1 - f_0)\sqrt{\frac{\pi}{\varepsilon_Q w}} [e^{-(\omega-\varepsilon_Q)^2/4\varepsilon_Q w} - e^{-(\omega+\varepsilon_Q)^2/4\varepsilon_Q w}] \tag{11.92}$$

Equation (11.91) predicts that the spectral function $S(\mathbf{Q}, \omega)$ is the difference of two Gaussians, peaked at $\omega = \pm\varepsilon_Q$ and with a width given by $2\sqrt{\varepsilon_Q w}$. There is also the delta function at $\omega = \pm\varepsilon_Q$ from the scattering from the condensate. The results appear as a single Gaussian whenever $\varepsilon_Q > \omega$, which is the usual experimental case.

There have been numerous measurements of $n_{\mathbf{k}}$, including those of Harling (1971), Rodriquez *et al.*, (1974), Svensson *et al.*, (1981), and Svensson and Sears (1986). The data of Woods and Sears (1977) is shown in Fig. 10.5. The solid line in Fig. 11.5(a) is a fit to their points taken at $T = 4.2$ K when the liquid ^{4}He is not superfluid. This line is a remarkably good fit to a Gaussian. The other set of points are taken at $T = 1.1$ K in the superfluid state. There is an obvious increase in the distribution $n_{\mathbf{k}}$ at small values of k, as shown by the difference spectra in Fig. 11.5(b). The increase at $k \approx 0$ has the width of their resolution function and is interpreted as being caused by the condensate fraction f_0. They estimate $f_0 = 6.9\% \pm 0.8\%$ from this data. If this value of f_0 is extrapolated to zero temperature by using the Bose–Einstein formula, one finds that $f_0(T = 0) = 10.8\% \pm 1.3\%$. The value $f_0 = 0.11$ is in good agreement with the theoretical estimates, with the earlier results of Harling, and with the analysis of Puff and Tenn. However, there is little justification for using the Bose–Einstein formula for extrapolating $f_0(T)$ to zero temperature except expediency, since no other formula is available. It also appears from the data of Woods and Sears that the values of $n_{\mathbf{k}}$ in the superfluid state are not as good a fit to a Gaussian as they were in the normal fluid. This feature is not understood.

Equation (11.86) for the sudden approximation is now derived. The first step begins with the definition of the spectral function $S(\mathbf{Q}, \omega)$. As in the case of the electron gas in Chapter 5, the spectral function is derived from the density–density correlation function. If

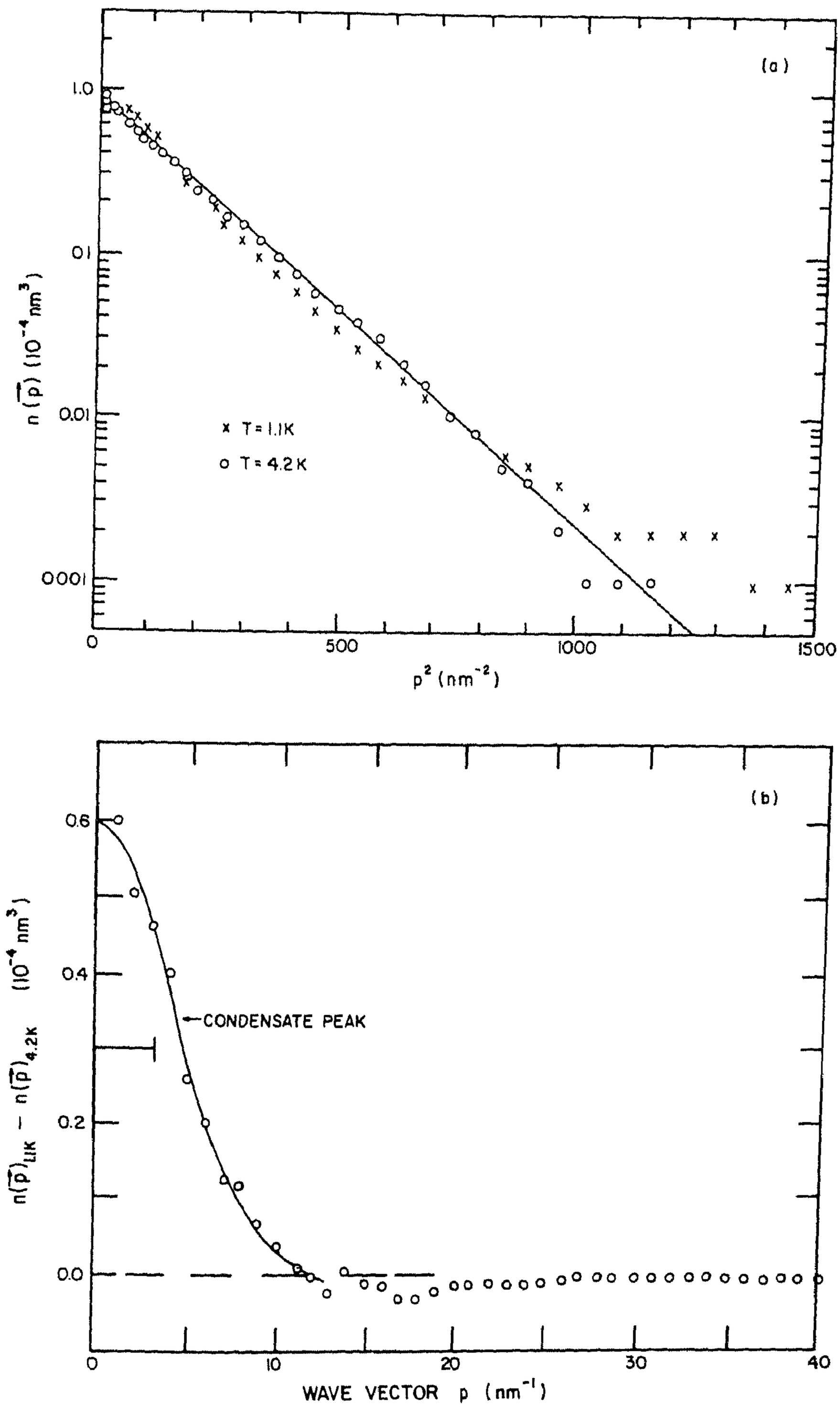


FIGURE 11.5 (a) Neutron scattering results for the momentum distribution in liquid ^{4}He at $T = 1.1$ and 4.2 K as a function of p^2 (b) The difference between the two distributions in part (a) shows a peak at small wave vector whose width equals the experimental resolution. The peak is interpreted as the condensate fraction. *Source:* Woods and Sears (1977) (used with permission).

this correlation function is called $\chi(\mathbf{Q}, i\omega_n)$ in the Matsubara representation, then $S(\mathbf{Q}, \omega)$ is -2 times the imaginary part of its retarded function:

$$\chi(\mathbf{Q}, i\omega_n) = -\frac{1}{N}\int_0^\beta d\tau e^{i\omega_n\tau}\langle T_\tau \rho(\mathbf{Q}, \tau)\rho(-\mathbf{Q}, 0)\rangle \tag{11.93}$$

$$S(\mathbf{Q}, \omega) = -2\,\mathrm{Im}[\chi_{\mathrm{ret}}(\mathbf{Q}, \omega)] \tag{11.94}$$

According to the result of Problem 16 in Chapter 3, $S(\mathbf{Q}, \omega)$ can written as

$$S(\mathbf{Q}, \omega) = \frac{1}{N}(1 - e^{-\beta\omega})\int_{-\infty}^{\infty} dt e^{i\omega t}\langle \rho(\mathbf{Q}, t)\rho(-\mathbf{Q}, 0)\rangle \tag{11.95}$$

The experiments are done with conditions of temperature and energy transfer $\hbar\omega$ such that $\beta\omega \gg 1$, and the second term in the parentheses $\exp(-\beta\omega)$ can be neglected. The space representation is used for the density operator $\rho(\mathbf{Q})$, which brings us to the expression

$$\rho(\mathbf{Q}) = \sum_j e^{i\mathbf{Q}\cdot\mathbf{R}_j} \tag{11.96}$$

$$S(\mathbf{Q}, \omega) = \frac{1}{N}\sum_{ij}\int_{-\infty}^{\infty} dt e^{i\omega t}\langle e^{i\mathbf{Q}\cdot\mathbf{R}_j(t)}e^{-i\mathbf{Q}\cdot\mathbf{R}_i(0)}\rangle \tag{11.97}$$

The correlation function is evaluated by determining how the particle motion $\mathbf{R}_j(t)$ proceeds in time. The two exponents $\mathbf{Q}\cdot\mathbf{R}_i$ and $\mathbf{Q}\cdot\mathbf{R}_j(t)$ cannot be combined, since they do not commute at different times. This point will be shown later.

The sudden approximation is obtained by solving the correlation function for small values of time t by expanding $\mathbf{R}_j(t)$ about point $t = 0$. The physical argument is that the scattering by energetic neutrons happens rapidly, so that only the short time response of the systems is applicable. Near $t = 0$, the coefficients in the Taylor series are found from the commutation relations:

$$\mathbf{R}_j(t) = \mathbf{R}_j(0) + t\left(\frac{d\mathbf{R}_j}{dt}\right)_0 + \frac{t^2}{2}\left(\frac{d^2\mathbf{R}_j}{dt^2}\right)_0 + O(t^3) \tag{11.98}$$

$$\frac{d\mathbf{R}_j}{dt} = i[H, \mathbf{R}_j] = \frac{\mathbf{P}_j}{m} \tag{11.99}$$

$$\frac{d^2\mathbf{R}_j}{dt^2} = \frac{i}{m}[H, \mathbf{P}_j] = -\frac{1}{m}\sum_n \nabla_j V(\mathbf{R}_j - \mathbf{R}_n) \equiv \frac{\mathbf{F}_j}{m} \tag{11.100}$$

The next step is to separate the several terms in the exponent into the product of several exponents. The Feynman theorem is used: $\exp(A + B) = \exp(A)\exp(B)\exp(C)$ with $B = i\mathbf{Q}\cdot\mathbf{R}_j$, $A = i(\mathbf{Q}/m)\cdot(t\mathbf{P}_j + t^2\mathbf{F}_j/2)$, and $C = -[A, B]/2 = -it\varepsilon_Q$. The next step is to separate the factor $it\mathbf{Q}\cdot\mathbf{P}_j/m$ from $it^2\mathbf{Q}\cdot\mathbf{F}_j/(2m)$. They do not commute, but such corrections terms contribute on the order $O(t^3)$, which are being neglected. The short time response is written as

$$S(\mathbf{Q}, \omega) = \frac{1}{N}\sum_{ij}\int_{-\infty}^{\infty} dt e^{it(\omega-\varepsilon_Q)}\langle e^{it\mathbf{Q}\cdot\mathbf{P}_j/m}e^{i(t^2/2m)\mathbf{Q}\cdot\mathbf{F}_j}e^{i\mathbf{Q}\cdot\mathbf{R}_j}e^{-i\mathbf{Q}\cdot\mathbf{R}_i}\rangle$$

The sudden approximation is achieved by making two important approximations to this expression. The first is to take only the term in the series with $i = j$. This choice is a valid approximation when $\mathbf{Q}$ is large, since then the terms with $i \neq j$ average to a small contribution. The second approximation is to neglect the term $it^2\mathbf{Q}\cdot\mathbf{F}_j/2m$. The quantity $\mathbf{F}_j$ is the

force on the particle at $\mathbf{R}_j$, which must average to zero. If this force is evaluated by a cumulant expansion, for example, the average $\langle \mathbf{F}_j \cdot \mathbf{Q} \rangle$ is zero, although the average $\langle (\mathbf{F}_j \cdot \mathbf{Q})^2 \rangle$ is not. This force contributes a term in the time expansion on the order of $O(t^4)$ rather than $O(t^2)$. It is neglected.

In the sudden approximation of high $\mathbf{Q}$ and high ω, the remaining expression to evaluate is:

$$\begin{aligned} S(\mathbf{Q}, \omega) &= \frac{1}{N} \sum_j \int_{-\infty}^{\infty} dt e^{it(\omega - \varepsilon_Q)} \langle e^{it\mathbf{Q} \cdot \mathbf{P}_j/m} \rangle \\ \langle e^{it\mathbf{Q} \cdot \mathbf{P}_j/m} \rangle &= \int d^3R_1 \cdots d^3R_N \Psi_0^*(\mathbf{R}_1 \cdots \mathbf{R}_N) e^{it\mathbf{Q} \cdot \mathbf{P}_j/m} \Psi_0(\mathbf{R}_1 \cdots \mathbf{R}_N) \end{aligned} \tag{11.101}$$

The quantity $\exp(it\mathbf{Q} \cdot \mathbf{P}_j/m)$ operates on the wave function to the right and displaces the position variable $\mathbf{R}_j$ by the increment $t\mathbf{Q}/m$:

$$e^{it\mathbf{Q} \cdot \mathbf{P}_j/m} \Psi_0(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N) = \Psi_0(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_j + t\mathbf{Q}/m, \ldots, \mathbf{R}_N)$$

The next step is to do all the position integrals except d^3R_j. The multiple integrals just produce the off-diagonal density matrix $\tilde{\rho}(\mathbf{R}_j, \mathbf{R}_j + t\mathbf{Q}/m)$. Since the off-diagonal density matrix is just a function of the difference of its arguments, the expression gives $\tilde{\rho}(\mathbf{R}_j, \mathbf{R}_j + t\mathbf{Q}/m) = R(t\mathbf{Q}/m)/v$:

$$S(\mathbf{Q}, \omega) = \frac{1}{N} \sum_j \int_{-\infty}^{\infty} dt e^{it(\omega - \varepsilon_Q)} R(t\mathbf{Q}/m) \tag{11.102}$$

$$S(\mathbf{Q}, \omega) = \int_{-\infty}^{\infty} dt e^{it(\omega - \varepsilon_Q)} R(t\mathbf{Q}/m) \tag{11.103}$$

The summation over j gives N, since all terms in the summation are alike. Earlier it was observed that $R(r)$ was approximately a Gaussian function of r. In this case, the integrand of the time integral is also a Gaussian. A Gaussian is found since the exponent is evaluated to order $O(t^2)$.

To derive to the standard formula for the sudden approximation, remember that $n_{\mathbf{k}}$ is the Fourier transform of $R(r)$:

$$R(\mathbf{r}) = \frac{1}{\rho_0} \int \frac{d^3k}{(2\pi)^3} n_{\mathbf{k}} e^{-i\mathbf{k} \cdot \mathbf{r}} \tag{11.104}$$

$$S(\mathbf{Q}, \omega) = \frac{1}{\rho_0} \int \frac{d^3k}{(2\pi)^3} n_{\mathbf{k}} \int_{-\infty}^{\infty} dt \exp\left[it\left(\omega - \varepsilon_Q - \frac{\mathbf{Q} \cdot \mathbf{k}}{m} \right) \right] \tag{11.105}$$

$$= \frac{2\pi}{\rho_0} \int \frac{d^3k}{(2\pi)^3} n_{\mathbf{k}} \delta\left(\omega - \varepsilon_Q - \frac{\mathbf{Q} \cdot \mathbf{k}}{m} \right) \tag{11.106}$$

which is the first term in (11.86). The second term in (11.86) comes from realizing that $S(\mathbf{Q}, \omega)$ is antisymmetric in ω.

Two important assumptions are made in the derivation of the sudden approximation. The first is that $\mathbf{Q}$ is large, and the second is that ω is large so that the Fourier transform involves only small values of t. The experiments are done such that both conditions are well met. Then the approximation is accurate and the experiments determine $n_{\mathbf{k}}$.

The sudden approximation has the appearance of treating the particles in the liquid as being free. The same result could be derived by assuming there was a gas of free particles with a distribution $n_{\mathbf{k}}$. When the neutron strikes a particle of momentum $\mathbf{k}$ and kinetic energy

Max - Planck - Institut
für Mikrostrukturphysik
Bibliothek

$\varepsilon_k = k^2/2m$, the particle is changed to wave vector $\mathbf{k}+\mathbf{Q}$ and kinetic energy $\varepsilon_{\mathbf{k}+\mathbf{q}} = (\mathbf{k}+\mathbf{Q})^2/2m$. Energy conservation demands that the energy transfer be $\omega + \varepsilon_{\mathbf{k}} = \varepsilon_{\mathbf{k}+\mathbf{Q}}$. The probability of having a momentum–energy transfer of $(\mathbf{Q}, \omega)$ is proportional to

$$\frac{1}{\rho_0}\int \frac{d^3k}{(2\pi)^3} n_{\mathbf{k}} \delta\left(\omega - \varepsilon_Q - \frac{\mathbf{Q}\cdot\mathbf{k}}{m}\right) \tag{11.107}$$

This equation is exactly the form of the sudden approximation. What happened to the potential energy? It is not really ignored. In the sudden approximation the essential physics is that when the neutron strikes the ^{4}He particle, its potential energy does not change immediately. The impulse will alter the particle velocity, and after a time duration the subsequent motion will alter its potential energy. However, by that time the neutron is gone and does not record this change in potential energy. The potential energy depends only on the position of the particles, and when the neutron strikes one particle and redirects its motion, the potential energy only changes later. The fast-moving neutron does not remain around to record this alteration in potential energy, so that the change in energy appears to involve only the kinetic energy component.

11.1.6. Bijl–Feynman theory

The excitation spectrum of ^{4}He is shown in Fig. 11.6 as determined by neutron scattering by Cowley and Woods (1971). These points are found as maxima in $S(\mathbf{Q}, \omega)$. For some

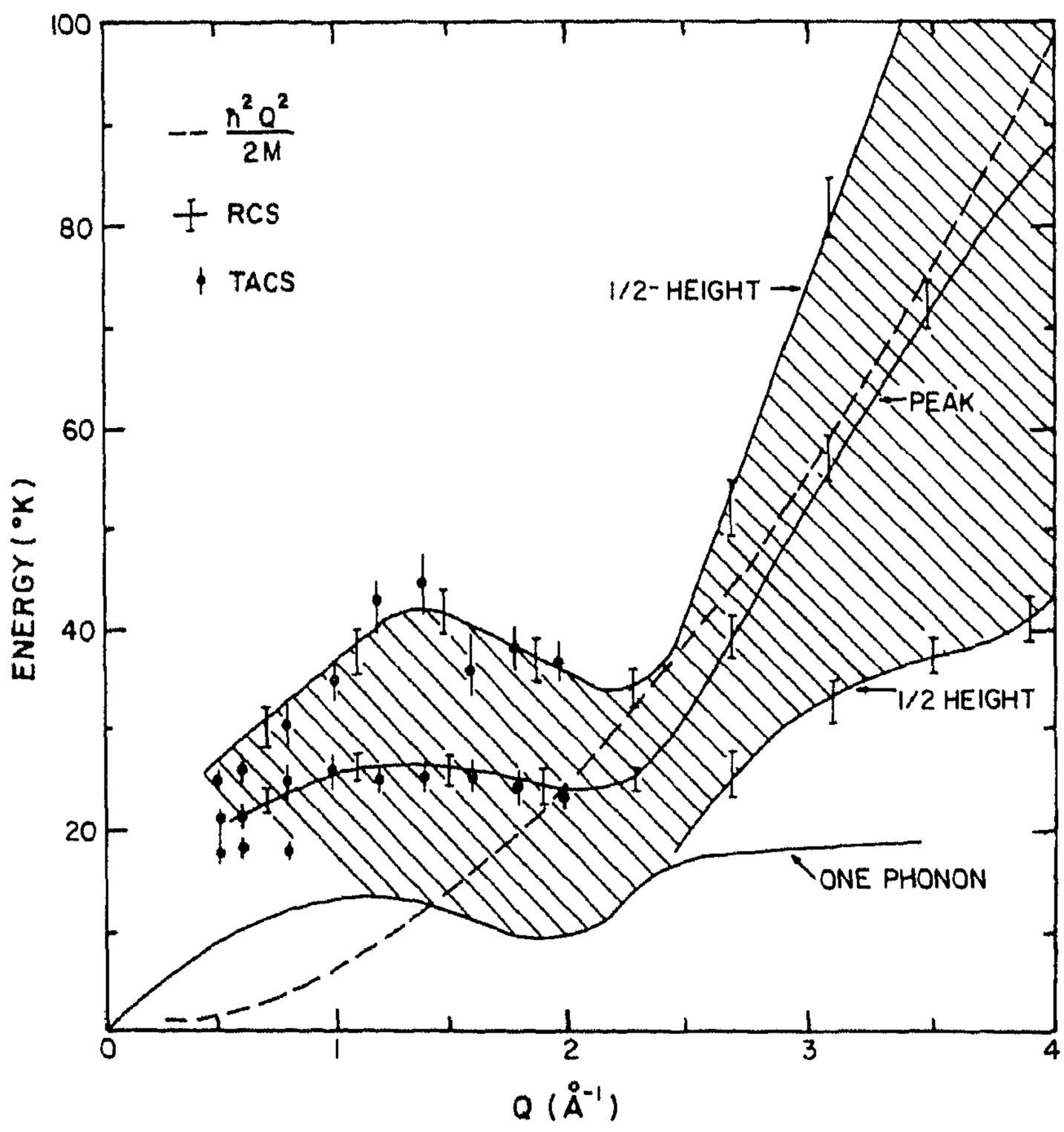


FIGURE 11.6 Neutron scattering results for the excitation spectrum of liquid ^{4}He at $T = 1.1$ K. The lower solid line is the one-phonon curve, while the dashed line is for the noninteracting particle. A broad second peak becomes the single-particle curve at large Q and the two-roton bound state at small Q. *Source:* Cowley and Woods (1971) (used with permission).

values of $\mathbf{Q}$ there are two values of ω, which indicates that $S(\mathbf{Q}, \omega)$ has two peaks. The dashed line shows the free-particle spectrum $\varepsilon_k = k^2/2m$. The excitation spectrum approaches this at large values of k, as shown by Harling (1971). At small values of k, the lower branch of the dispersion curve becomes linear in k. Small wave vector is the phonon region of sound waves, and the slope is the sound velocity, 240 m/s. The lower dispersion curve has a second minimum which is called the roton. This point has a high density of states and so has a large weight when averaging over the density of excitation states. Indeed, the necessity of this minimum was deduced by Landau (1941, 1947) on the basis of thermodynamic data. The roton parameters are $k_0 = 1.91$ Å^{-1} and $\Delta = 8.68$ K. There is a two-roton bound state at 2Δ which forms the lower limit of the upper dispersion curve (Ruvalds and Zawadowski, 1970) Zawadowski *et al.* (1972). The temperature dependence of the excitations was measured by Stirling and Glyde (1990).

This experimental curve forms the basis for the low-temperature and superfluid properties of ^{4}He. It explains all macroscopic behavior, based on the Landau two-fluid model. Once the excitation spectrum is derived, superfluidity can be explained. Deriving the excitation spectrum is the objective of this section.

The successful method of finding the excitation spectrum of liquid ^{4}He was popularized by Feynman (1954). He suggested that the excitation spectrum is given by the formula $\varepsilon_k/S(k)$, where $S(k)$ is the liquid structure factor. This formula is qualitatively right but not quantitatively accurate. However, its importance is that it led to the formalism which does produce the right result. The same formula $\varepsilon_k/S(k)$ was obtained much earlier by Bijl (1940), whose suggestion attracted little notice since neither $S(k)$ nor ω_k were known in those early days. The current usage is to call this expression the Bijl–Feynman formula:

$$\omega_0(k) = \frac{\varepsilon_k}{S(k)} \tag{11.108}$$

This result is called $\omega_0(k)$ and is derived next.

The goal is to construct a wave function $\Psi_{\mathbf{k}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ which describes an excitation in the system with a wave vector $\mathbf{k}$. For the boson system of liquid ^{4}He, this excitation wave function must still have the property of being symmetric under the interchange of any two-particle coordinates. For long-wavelength excitations, the wavefunction will still have the same kind of short-range correlations which were found for the ground state wave function. It seems reasonable to assume that the excited state wave function is only a slight perturbation upon the ground state wave function. A suitable wave function is obtained by writing the excited state as the product of the ground state wave function and another symmetric function $\Lambda_{\mathbf{k}}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ of all the positions:

$$\Psi_{\mathbf{k}}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \Lambda_{\mathbf{k}}(\mathbf{r}_1, \ldots, \mathbf{r}_N)\Psi_0(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{11.109}$$

The idea is simple enough. The stipulation on $\Lambda_{\mathbf{k}}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is that it is symmetric in the coordinates and contains the wave vector $\mathbf{k}$. Two possible choices are

$$\Lambda_{\mathbf{k}}^{(1)}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = L_{\mathbf{k}} \sum_l e^{i\mathbf{k}\cdot\mathbf{r}_l} = L_{\mathbf{k}}\rho(\mathbf{k}) \tag{11.110}$$

$$\Lambda_{\mathbf{k}}^{(2)}(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \exp\left(i\mathbf{k}\cdot \sum_l \mathbf{r}_l\right) \tag{11.111}$$

Perhaps the reader can think of others. Each choice of $\Lambda_{\mathbf{k}}$ corresponds to some kind of excited state of the system. What is the physics behind each one? It is helpful to use operators such as the total momentum of the system:

$$\mathbf{P} = \frac{\hbar}{i}\sum_j \nabla_j \tag{11.112}$$

For example, the total momentum on $\Psi^{(2)} = \Lambda_{\mathbf{k}}^{(2)}\Psi_0$ gives two terms, since $\mathbf{P}$ can act upon either $\Lambda_{\mathbf{k}}^{(2)}$, or Ψ_0. However, the average momentum in the ground state must be zero, so this can be evaluated to give

$$\begin{aligned}\langle\Psi_{\mathbf{k}}^{(2)}|\mathbf{P}|\Psi_{\mathbf{k}}^{(2)}\rangle &= \int d^3r_1\cdots d^3r_N\Psi_0^*\Lambda_{\mathbf{k}}^{(2)*}[\Psi_0\mathbf{P}\Lambda_{\mathbf{k}}^{(2)} + \Lambda_{\mathbf{k}}^{(2)}\mathbf{P}\Psi_0]\\ &= \sum_j \hbar\mathbf{k}\int d^3r_1\cdots d^3r_N\Psi_0^*\Lambda_{\mathbf{k}}^{(2)*}\Lambda_{\mathbf{k}}^{(2)}\Psi_0 = N\hbar\mathbf{k}\end{aligned}$$

since $\Lambda^{(2)*}\Lambda^{(2)} = 1$. The total momentum of a system of N particles is just $N\mathbf{k}$, which corresponds to the uniform flow of the entire fluid with velocity $\mathbf{k}/m$. That is the correct interpretation of the excitation spectrum represented by the choice $\Psi^{(2)} = \Lambda_{\mathbf{k}}^{(2)}\Psi_0$. Another check on this interpretation is to find the energy by evaluating the Hamiltonian:

$$E = \frac{1}{N}\langle\Psi_{\mathbf{k}}^{(2)}|H|\Psi_{\mathbf{k}}^{(2)}\rangle \tag{11.113}$$

$$= -\frac{\hbar^2}{2mN}\sum_j\int d^3r_1\cdots d^3r_N\Psi_0^*\Lambda_{\mathbf{k}}^{(2)*}\left[\Psi_0\nabla_j^2\Lambda_{\mathbf{k}}^{(2)} + 2\nabla_j\Lambda_{\mathbf{k}}^{(2)}\cdot\nabla_j\Psi_0 + \Lambda_{\mathbf{k}}^{(2)}\nabla_j^2\Psi_0\right]$$

$$+\frac{1}{N}\int d^3r_1\cdots d^3r_N|\Psi_0|^2\sum_{ij}V(r_{ij}) \tag{11.114}$$

$$E = \varepsilon_{\mathbf{k}} + \frac{E_0}{N} \tag{11.115}$$

The average energy per particle is $\varepsilon_k = k^2/2m$ plus the ground state value E_0/N. The cross term $2\nabla_j\Lambda_{\mathbf{k}}^{(2)}\cdot\nabla_j\Psi_0$ in the preceding expression averages to zero, since it contains the average momentum $\int\Psi_0^*\nabla\Psi_0$ in the ground state of the stationary liquid.

The other excited state wave function $\Psi_{\mathbf{k}}^{(1)} = \Lambda_{\mathbf{k}}^{(1)}\Psi_0$ is the important one for the excitation spectrum of the stationary liquid. The superscript (1) will be dropped, and it will be called just $\Psi_{\mathbf{k}} = \Lambda_{\mathbf{k}}\Psi_0 = L_{\mathbf{k}}\rho(\mathbf{k})\Psi_0$. This term will be investigated systematically. The first step is to find the normalization constant $L_{\mathbf{k}}$ for the wave function, which is done by setting the normalization integral to unity:

$$1 = \langle\Psi_{\mathbf{k}}|\Psi_{\mathbf{k}}\rangle = L_{\mathbf{k}}^2\sum_{jl}\int d^3r_1\cdots d^3r_N|\Psi_0|^2\exp[i\mathbf{k}\cdot(\mathbf{r}_l - \mathbf{r}_j)] \tag{11.116}$$

The terms in the double summation with $l = j$ just give unity since the wave vector dependence drops out. Special care must be taken whenever $l \neq j$. Then the space integrals can each be evaluated, at least formally, except those with $\mathbf{r}_j$ and $\mathbf{r}_l$. These integrals over $d^3r_1\cdots d^3r_N$ just produce the pair distribution function $\rho_2(\mathbf{r}_j, \mathbf{r}_l) = v^{-2}g(\mathbf{r}_j - \mathbf{r}_l)$:

$$1 = L_{\mathbf{k}}^2\left[\sum_{j=l}1 + \sum_{j\neq l}\frac{1}{v^2}\int d^3r_jd^3r_lg(\mathbf{r}_j - \mathbf{r}_l)e^{i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)}\right] \tag{11.117}$$

The summation for each particle l is the same, since all particles are identical. This summation is eliminated and instead multiply by N:

$$1 = NL_{\mathbf{k}}^2\left[1 + \sum_{j\neq 1}\frac{1}{v^2}\int d^3r_j d^3r_1 g(\mathbf{r}_j - \mathbf{r}_1)e^{i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_1)}\right] \tag{11.118}$$

Each term in the summation over j gives the same result, so just do one term and multiply by $N - 1 \approx N$. Also, change integration variables to relative coordinates $d^3r_j d^3r_1 = vd^3r$, so the integral is over $g(r)$. In fact, this expression is precisely the definition of $S(\mathbf{k})$ which is given earlier in Sec. 1.6:

$$1 = NL_{\mathbf{k}}^2\left(1 + \rho\int d^3r g(r)e^{i\mathbf{k}\cdot\mathbf{r}}\right) = NS(\mathbf{k})L_{\mathbf{k}}^2 \tag{11.119}$$

$$L_{\mathbf{k}} = \frac{1}{\sqrt{NS(\mathbf{k})}} \tag{11.120}$$

Equation (11.120) provides the rigorous evaluation of the normalization constant $L_{\mathbf{k}}$. It is useful to call the structure factor $S(\mathbf{k})$ although it is only a function of $|\mathbf{k}| = k$ in the liquid.

Now that the wave function is normalized properly, it is possible to investigate its properties by taking the expectation of different operators. The first is the momentum operator $\mathbf{P}$ given in (11.112):

$$\langle\mathbf{P}\rangle = \langle\Psi_{\mathbf{k}}|\mathbf{P}|\Psi_{\mathbf{k}}\rangle \tag{11.121}$$

First examine a single gradiant of position ∇_j acting upon the wave function $\Psi_{\mathbf{k}}$. Since the function $\Lambda_{\mathbf{k}}$ is a summation over different coordinates, the one term with $l = j$ will be different from the others:

$$\nabla_j\Psi_{\mathbf{k}}(\mathbf{r}_1, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_N) = L_{\mathbf{k}}\sum_l \nabla_j(e^{i\mathbf{k}\cdot\mathbf{r}_l}\Psi_0) \tag{11.122}$$

$$= L_{\mathbf{k}}\left[\nabla_j(e^{i\mathbf{k}\cdot\mathbf{r}_j}\Psi_0) + \sum_{l\neq j} e^{i\mathbf{k}\cdot\mathbf{r}_l}\nabla_j\Psi_0\right]$$

$$= i\mathbf{k}L_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}_j}\Psi_0 + \Lambda_{\mathbf{k}}\nabla_j\Psi_0 \tag{11.123}$$

The momentum has a simple result

$$\mathbf{P}\Psi_{\mathbf{k}} = \hbar\mathbf{k}\Psi_{\mathbf{k}} + \Lambda_{\mathbf{k}}\mathbf{P}\Psi_0 \tag{11.124}$$

$$\langle\mathbf{P}\rangle = \hbar\mathbf{k} + \langle\Psi_{\mathbf{k}}|\Lambda_{\mathbf{k}}\mathbf{P}\Psi_0\rangle \tag{11.125}$$

$$= \hbar\mathbf{k} + \langle\Psi_0|\mathbf{P}|\Psi_0\rangle = \hbar\mathbf{k} \tag{11.126}$$

The first term gives $\hbar\mathbf{k}$. The second term averages to zero because the operators $\Lambda_{\mathbf{k}}^\dagger\Lambda_{\mathbf{k}}$ average to unity, and then $\langle\Psi_0|\mathbf{P}|\Psi_0\rangle$ is the momentum in the ground state, which is zero. The momentum in state $\Psi_{\mathbf{k}}$ is $\hbar\mathbf{k}$. The momentum operator commutes with the original Hamiltonian, so that states can be described by the quantum number of momentum $\hbar\mathbf{k}$.

The next step is to calculate the energy of this excitation whose wave vector is $\mathbf{k}$. The energy is found by taking the expectation value of the Hamiltonian operator between these states:

$$E_{\mathbf{k}} = \langle\Psi_{\mathbf{k}}|H|\Psi_{\mathbf{k}}\rangle = L_{\mathbf{k}}^2\langle\Psi_0\rho(-\mathbf{k})H\rho(\mathbf{k})\Psi_0\rangle \tag{11.127}$$

It is desirable to commute the Hamiltonian operator to the right, so that it operates on the ground state Ψ_0. The ground state is an exact eigenstate of the Hamiltonian $H\psi_0 = E_0\Psi_0$ and is the one with the lowest eigenvalue E_0.

$$H\rho(\mathbf{k}) = H\rho(\mathbf{k}) - \rho(\mathbf{k})H + \rho(\mathbf{k})H = [H, \rho(\mathbf{k})] + \rho(\mathbf{k})H \tag{11.128}$$

$$E_\mathbf{k} = L_\mathbf{k}^2\langle\Psi_0\rho(-\mathbf{k})[H, \rho(\mathbf{k})]\Psi_0\rangle + L_\mathbf{k}^2\langle\Psi_0\rho(-\mathbf{k})\rho(\mathbf{k})H\Psi_0\rangle \tag{11.129}$$

The second term on the right gives $E_0L_\mathbf{k}^2\langle\Psi_0\rho(-\mathbf{k})\rho(\mathbf{k})\Psi_0\rangle = E_0$. The first term on the right contains the commutator of H with the density operator $\rho(\mathbf{k})$. Only the kinetic energy term fails to commute with $\rho(\mathbf{k})$ so

$$[H, \rho(\mathbf{k})] = -\frac{\hbar^2}{2m}\sum_{ij}[\nabla_i^2, e^{i\mathbf{k}\cdot\mathbf{r}_j}] \tag{11.130}$$

$$= \frac{\hbar^2}{2m}\sum_j e^{i\mathbf{k}\cdot\mathbf{r}_j}[k^2 - 2i\mathbf{k}\cdot\nabla_j] \tag{11.131}$$

$$\langle\Psi_0\rho(-\mathbf{k})[H, \rho(\mathbf{k})]\Psi_0\rangle = \varepsilon_k\langle\Psi_0\rho(-\mathbf{k})\rho(\mathbf{k})\Psi_0\rangle \tag{11.132}$$

$$- i\frac{\hbar^2}{m}\sum_j\langle\Psi_0\rho(-\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}_j}\mathbf{k}\cdot\nabla_j\Psi_0\rangle \tag{11.133}$$

The second term on the right must be treated carefully. Since the ground state wave function is real, we can replace $\Psi_0\nabla_j\Psi_0$ by $\nabla_j(\Psi_0)^2/2$ and then integrate by parts, so that this term is transformed to

$$\langle\Psi_0\rho(-\mathbf{k})[H, \rho(\mathbf{k})]\Psi_0\rangle$$

$$= \varepsilon_k\langle\Psi_0\rho(-\mathbf{k})\rho(\mathbf{k})\Psi_0\rangle + i\frac{\hbar^2}{2m}\sum_j\langle\rho_N\mathbf{k}\cdot\nabla_j[\rho(-\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}_j}]\rangle \tag{11.134}$$

$$= \sum_{jl}\left\langle\rho_N\left[\varepsilon_k + \frac{i\hbar^2}{2m}\mathbf{k}\cdot\nabla_j\right]e^{i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)}\right\rangle \tag{11.135}$$

The gradient operator in the last term will produce a factor $i\mathbf{k}$ in each term except the one where $l = j$. The two terms will cancel term by term except the one where $l = j$, which exists only in the first term. The result is

$$\langle\Psi_0\rho(-\mathbf{k})[H, \rho(\mathbf{k})]\Psi_0\rangle = \varepsilon_k\sum_j 1 = N\varepsilon_k \tag{11.136}$$

Since $L_\mathbf{k}^2 = 1/NS(\mathbf{k})$, the first term in (11.129) for $E_\mathbf{k}$ is $\varepsilon_\mathbf{k}/S(\mathbf{k})$, and the second is E_0. The energy in the system for the excitation $\mathbf{k}$ is

$$E_\mathbf{k} = \langle\Psi_\mathbf{k}|H|\Psi_\mathbf{k}\rangle = E_0 + \omega_0(\mathbf{k}) \tag{11.137}$$

$$\omega_0(\mathbf{k}) = \frac{\varepsilon_\mathbf{k}}{S(\mathbf{k})} \tag{11.138}$$

The excitation energy is defined as $\omega_0(\mathbf{k})$, since it is the energy above the ground state E_0. The result is the Bijl–Feynman excitation energy.

The solid line in Fig. 11.7 shows $\omega_0(\mathbf{k})$ and is compared with the actual excitation spectrum of liquid ^{4}He. The two curves have exactly the same shape, but the Bijl–Feynman spectrum is about a factor of two larger. The roton minimum in the Bijl–Feynman dispersion curve is caused by the peak in $S(\mathbf{k})$ at $k \approx 2$ Å. At small values of k, $\omega_0(\mathbf{k})$ goes to zero

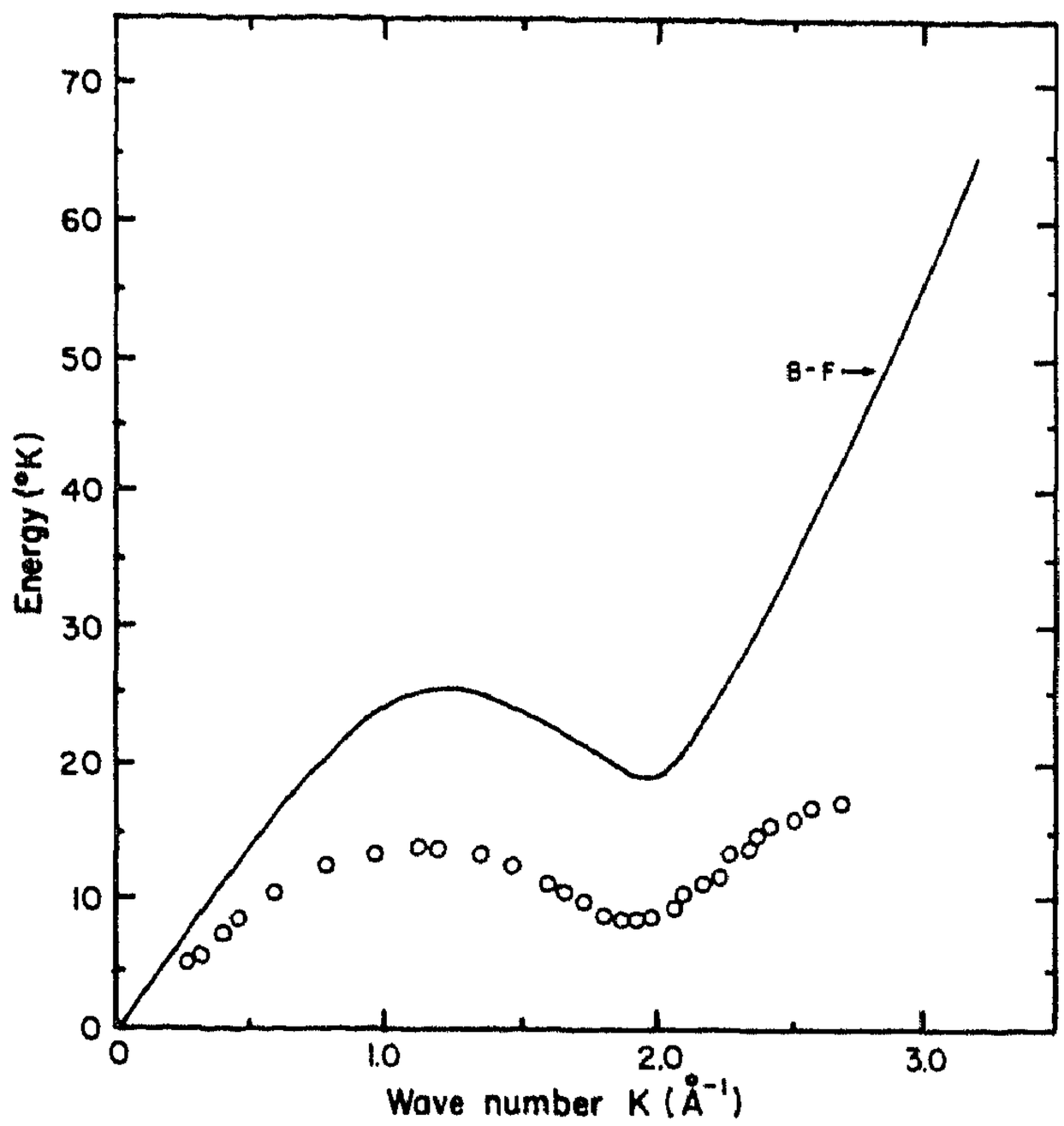


FIGURE 11.7 The excitation spectrum of liquid ^{4}He. The solid line is the Bijl–Feynman theory, compared with the experimental data as the points.

linearly with k, and the slope is the speed of sound. In the Bijl–Feynman theory, this happens because ε_k vanishes as k^2 while $S(\mathbf{k})$ vanishes as k, so their ratio does vanish as k. One cannot use experimental values of $S(\mathbf{k})$ in this region, since the experimental values of $S(\mathbf{k})$ have an additional contribution from thermal fluctuations which provides a small constant to $S(\mathbf{k})$ in the limit where $k \to 0$. This constant, which is proportional to $k_B T$ times the compressibility, is easily subtracted from experimental $S(\mathbf{k})$ to obtain the $S(\mathbf{k})$ one desires to use in the Bijl–Feynman formula.

The agreement between theory $\omega_0(\mathbf{k})$ and experiment $\omega(\mathbf{k})$ is not satisfactory. The theory must press onward, and one must look for additional interactions or mechanisms which will improve the comparison. The interactions between these excitations must be included. Next consider a system with multiple excitations, where each is obtained by multiplying the ground state by the factor $\Lambda_{\mathbf{k}} = L_{\mathbf{k}}\rho(\mathbf{k})$:

$$\begin{aligned}
\Psi_{\mathbf{k}} &= \Lambda_{\mathbf{k}}\Psi_0 \equiv |\mathbf{k}) \\
\Psi_{\mathbf{k}_1,\mathbf{k}_2} &= \Lambda_{\mathbf{k}_1}\Lambda_{\mathbf{k}_2}\Psi_0 \equiv |\mathbf{k}_1, \mathbf{k}_2) \\
\Psi_{\mathbf{k}_1,\ldots,\mathbf{k}_n} &= \Lambda_{\mathbf{k}_1}\cdots\Lambda_{\mathbf{k}_n}\Psi_0 \equiv |\mathbf{k}_1, \ldots, \mathbf{k}_n)
\end{aligned} \tag{11.139}$$

The $\Lambda_{\mathbf{k}}$ are scalar functions, so it does not matter in which order they are arranged. These states are examined carefully in the next section to see whether the different excitations are independent or whether they interact.

11.1.7. Improved Excitation Spectra

The method of Feenberg (1969) will be described, which is based upon the states in (11.139) which contain multiple excitations. They are examined to test their orthogonality and their interaction. The state of one excitation $|\mathbf{k})$ has an interaction with the state of two

excitations $|\mathbf{k}', \mathbf{k}-\mathbf{k}')$. They contribute self-energy effects for the state of one interaction. These self-energy effects are just what is necessary to give a good spectrum. The final result will be $\omega(\mathbf{k}) = \omega_0(\mathbf{k}) + \Sigma(\mathbf{k})$, where $\Sigma(\mathbf{k})$ is the self-energy of the excitation $\mathbf{k}$. The method of calculating this self-energy is only outlined; a full description is in Feenberg's book.

The reader should be aware that the present procedure is dramatically different from any used elsewhere in this book. All the other problems have been based on the interaction representation, where the Hamiltonian is written as $H = H_0 + V$, where V is a well defined perturbation. The present approach does not do this at all. Instead, a set of states such as (11.139) is written down as an ansatz. These states cannot be readily identified with any part of the Hamiltonian and are not orthogonal eigenstates of any operator which is obvious. Nor is it straightforward to do perturbation theory, since there is no V in which to expand. These problems can all be overcome, but the reader should be alerted that the present derivation is quite different from previous ones.

Feenberg and his associates established the following properties for these excitation states:

$$(\mathbf{kp}|\mathbf{kp}) = 1, \qquad \mathbf{k} \neq \mathbf{p} \tag{11.140}$$

$$(\mathbf{kk}|\mathbf{kk}) = 2 \tag{11.141}$$

$$(\mathbf{kp}|H|\mathbf{kp}) = E_0 + \omega_0(\mathbf{k}) + \omega_0(\mathbf{p}) \tag{11.142}$$

These properties are exactly those for operators $\Lambda_\mathbf{k}$ which are were treated as boson creation operators $a_\mathbf{k}^\dagger$. For example, the state with two excitations in $\mathbf{k}$ is

$$|\mathbf{k}) = a_\mathbf{k}^\dagger|0\rangle = |1_\mathbf{k}\rangle \tag{11.143}$$

$$|\mathbf{kk}) = a_\mathbf{k}^\dagger a_\mathbf{k}^\dagger|0\rangle = a_\mathbf{k}^\dagger|1_\mathbf{k}\rangle = \sqrt{2}|2_\mathbf{k}\rangle \tag{11.144}$$

where the $\sqrt{2}$ comes from the usual raising operators for bosons. This factor explains the additional factor of 2 in (11.141). The state $|\mathbf{k}_1\mathbf{k}_2)$ has an excitation energy given by $\omega_0(\mathbf{k}_1) + \omega_0(\mathbf{k}_2)$ and seems to be composed of states with independent boson excitations.

However, this free-boson analogy should not be pushed too literally. For example, states with different numbers of bosons but with the same total momentum are orthogonal only to order $O(1/\sqrt{N})$. For example,

$$(\mathbf{k}_1, \mathbf{k}_2|\mathbf{k}_1 + \mathbf{k}_2) = \left[\frac{S(\mathbf{k}_1)S(\mathbf{k}_2)S(\mathbf{k}_1 + \mathbf{k}_2)}{N}\right]^{1/2} \tag{11.145}$$

$$(\mathbf{k}_1, \mathbf{k}_2|H|\mathbf{k}_1 + \mathbf{k}_2) = O\left(\frac{1}{\sqrt{N}}\right) \tag{11.146}$$

The last two matrix elements are important for calculating the self-energy of the excitations. There exists a matrix element for one excitation of wave vector $\mathbf{k}$ going to two states with $\mathbf{k}'$ and $\mathbf{k}-\mathbf{k}'$. The self-energy diagram is shown in Fig. 11.8. The excitations are represented by wavy lines, and the one excitation has a self-energy from making two excitations. The same Feynman diagram is calculated for the self-energy of a phonon from anharmonic interactions.

The matrix elements in (11.146) show that the excitations are not truly independent. Anharmonic processes contribute to the self-energy. It is derived in the following fashion. Since there is no obvious interaction V in the present problem, instead use $H - E$. The basic matric element is $(\mathbf{p}, \mathbf{k}-\mathbf{p}|(H-E)|\mathbf{k})$ which uses both results in (11.146). From second order Brillouin–Wigner perturbation theory, the self-energy is

$$\Sigma(\mathbf{k}, E) = \tfrac{1}{2}\sum_\mathbf{p} \frac{|(\mathbf{p}, \mathbf{k}-\mathbf{p}|(H-E)|\mathbf{k})|^2}{E - E_0 - \omega_0(\mathbf{p}) - \omega_0(\mathbf{k}-\mathbf{p})} \tag{11.147}$$

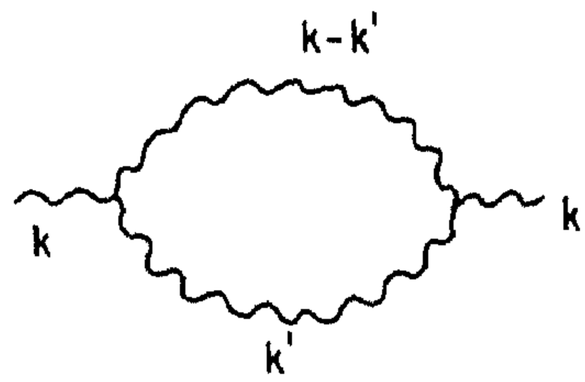


FIGURE 11.8

A self-consistent equation is derived for $\omega(\mathbf{k}) = \omega_0(\mathbf{k}) + \Sigma(\mathbf{k})$. Note that E appears in the matrix element and the energy denominator. Feenberg and his collaborators took the quantity $(H - E) \approx (H - E_0 - \omega_0(\mathbf{k}))$ in the matrix element, and found only an approximate eigenvalue from their results. The details of calculating the matrix element of H between the excitation states is tedious, since it involves an evaluation of three-particle correlation functions for the liquid. Their numerical results are shown in Fig. 11.9. as the dashed line. They compare quite favorably with the experimental spectrum as found by neutron scattering. The result is a dramatic improvement over the Bijl–Feynman curve, which is marked as B–F. Probably the first calculation of this kind was done by Feynman and Cohen (1956, 1957), who did a variational calculation of the backflow around the excitation. That is exactly the same physics as the admixing of the double excitation states, so the two methods of calculation are really the same but dressed in a slightly different language.

The theory of liquid ^{4}He is remarkably successful. Starting from the potential function $V(r)$ between atoms and other basic numbers such as the particle mass and density, the theory has calculated most of the features of the liquid with great success. By using correlated basis functions, the ground state properties such as $g(r)$, $S(\mathbf{k})$, $n_\mathbf{k}$, condensate fraction f_0, and off-

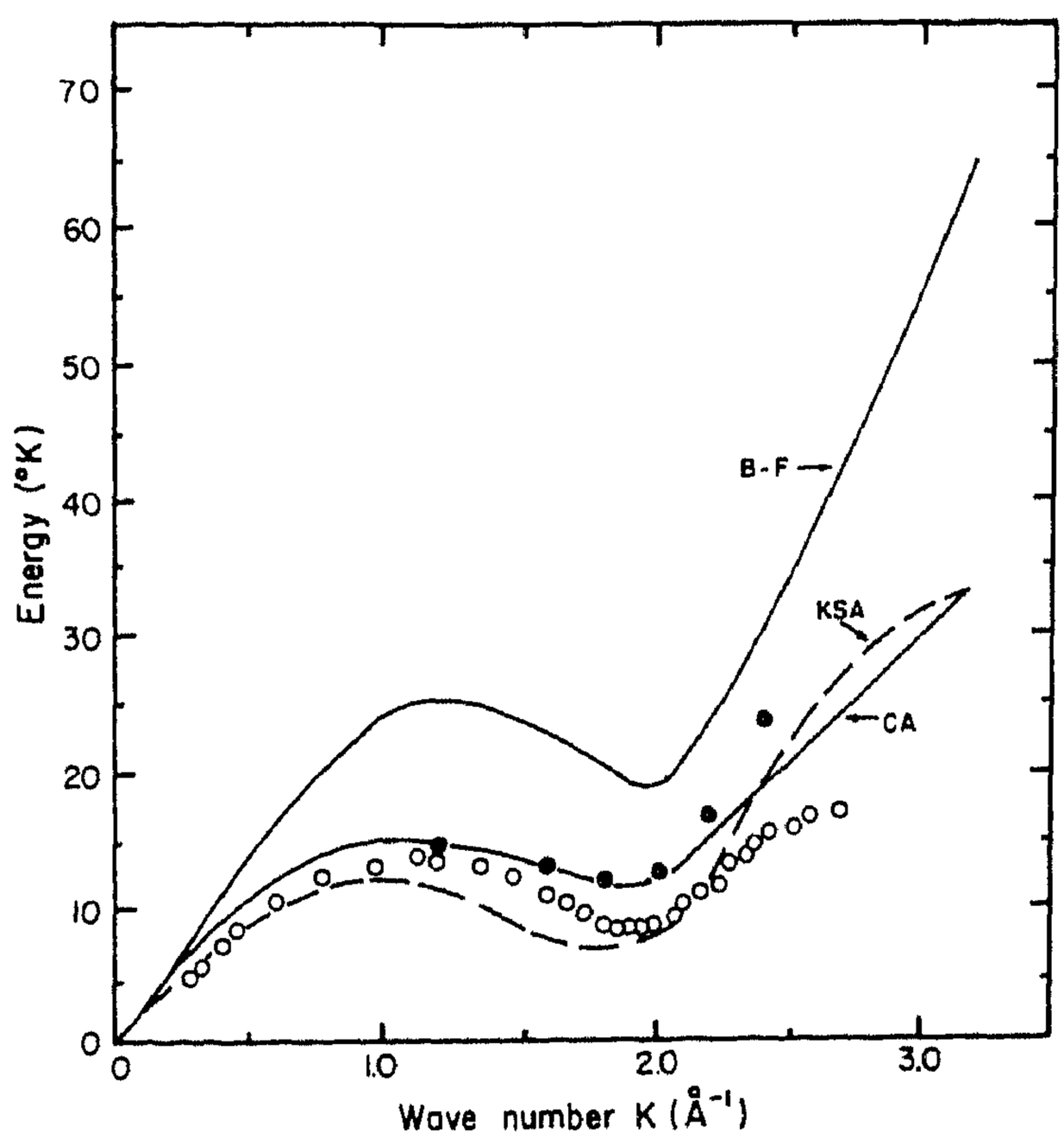


FIGURE 11.9 The excitation spectrum of liquid ^{4}He. The points are the experimental result for the superfluid. The other lines are various theories, including Bijl–Feynman (B–F) and backflow theories which use the Kirkwood superposition approximation (KSA) or convolution approximation (CA). Solid points are the Feynman–Cohen theory. *Source:* Feenberg (1969) (used with permission).

diagonal density matrix $R(r)$ are obtained. Most of these quantities, or their Fourier transforms, can be compared favorably with experimental results. Similarly, the excitation spectrum can also be calculated. After some work, a decent curve for $\omega(\mathbf{k})$ is now found which compares well with the experimental results. All of the features of liquid ^{4}He have been derived starting from a microscopic theory. The excitation spectrum $\omega(\mathbf{k})$ is all that is required to find the macroscopic properties, such as the superfluid density in the two-fluid model, and the superfluid flow properties. The macroscopic theory of superfluidity in ^{4}He can be derived, step by step, from a knowledge of only $V(r)$ plus some fundamental constants and masses. This achievement is remarkable for such a strongly interacting and highly correlated system.

11.1.8. Superfluidity

This section explains why the liquid is superfluid, and shows how to compute its superfluid properties. The explanation was first provided by Landau (1941, 1947), whose derivation is given below.

For temperatures below T_λ in the superfluid state, the fluid density can be viewed as consisting of a normal ρ_N and superfluid ρ_S component. The sum of these two is the normal density $\rho = \rho_N + \rho_S$. The two components ρ_N and ρ_S vary considerably with temperature where $\rho_S(T_\lambda) = 0$ and $\rho_N(T = 0) = 0$. The actual variation with temperature of ρ_N/ρ is shown in Fig. 11.10. In this two-fluid model, each component of the fluid, normal and superfluid, carries its own momentum, energy, etc. The normal component is given by the following expression for $T < T_\lambda$

$$\rho_N = \frac{\hbar^2}{3m}\int \frac{d^3q}{(2\pi)^3}\, q^2 \left\{ -\frac{d}{d\omega(q)} n_B[\omega(q)] \right\} \tag{11.148}$$

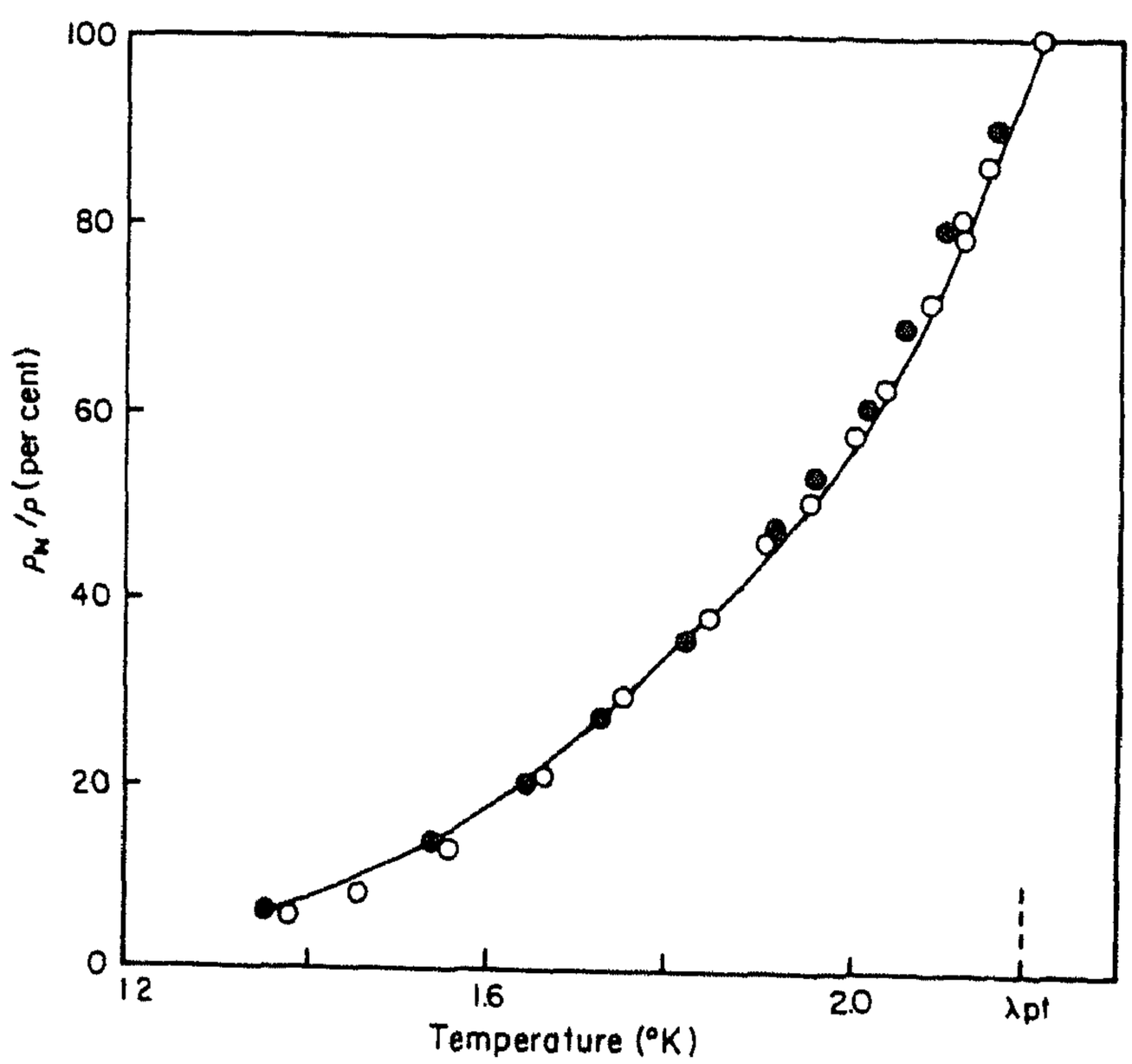


FIGURE 11.10 The density of the normal fluid ρ_N as a function of temperature: ○, Derived from oscillating disk experiments; •, from the velocity of second sound. *Source:* Lifshits and Andronikashivile (1959).

which is explained later. The superfluid component is just $\rho_S = \rho - \rho_N$ and has nothing at all to do with the condensate fraction f_0. If one takes the measured excitation spectrum $\omega(q)$ and uses it to theoretically calculate $\rho_N(T)$, one obtains the solid line in Fig. 11.10. This calculation is in remarkably good agreement with a number of ways to measure $\rho_N(T)$ such as the viscous drag on rotating plates, etc. Indeed, Landau was able to deduce a number of properties of $\omega(q)$ before it was measured directly from neutron scattering by fitting Eq. (11.148) to the measured component of normal density. In particular, the curves for $\rho_N(T)$ show an activation energy, which is now associated with the roton minimum Δ. Landau was able to deduce the necessity for this roton minimum, as well as determine the speed of sound, and predicted quite well the entire shape of the excitation spectrum, including phonon and roton parts. His was a remarkable theoretical achievement.

The Landau argument may be understood by considering a Gedanken experiment, which has the liquid helium flowing down a pipe with a uniform velocity $\mathbf{v}$. This velocity is small, say 1 cm/s, which is much smaller than the velocities in the zero-point motion of the fluid. The total momentum of the particle flow is calculated, and it is *not* just $Nm\mathbf{v}$, where N is the number of particles.

First construct a wave function which describes the uniformly moving liquid, which uses the operator $\Lambda_{\mathbf{k}}^{(2)}$ for uniform flow in (11.110). Label this wave function as $\Psi_{\mathbf{v}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ to denote the state of uniform flow. The energy and momentum of this wave function were calculated earlier:

$$\Psi_{\mathbf{v}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \exp\left(im\mathbf{v}\cdot \sum_j \mathbf{r}_j\right)\Psi_0(\mathbf{r}_1 \cdots \mathbf{r}_N) \tag{11.149}$$

$$\langle \Psi_{\mathbf{v}}|\mathbf{P}|\Psi_{\mathbf{v}}\rangle = mN\mathbf{v} \tag{11.150}$$

$$\langle \Psi_{\mathbf{v}}|H|\Psi_{\mathbf{v}}\rangle = E_0 + N\frac{mv^2}{2} \tag{11.151}$$

The next step is to consider the excitation spectrum of this moving fluid. The multiplication of $\Psi_{\mathbf{v}}$ by $\Lambda_{\mathbf{k}} = L_{\mathbf{k}}\rho(\mathbf{k})$ produces an excitation with momentum $\hbar\mathbf{k}$ and excitation energy $\omega_0(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v}$, as shown by

$$\phi_{\mathbf{v}}(\mathbf{k}) = L_{\mathbf{k}}\rho(\mathbf{k})\Psi_{\mathbf{v}} \tag{11.152}$$

$$\langle \phi_{\mathbf{v}}(\mathbf{k})|\mathbf{P}|\phi_{\mathbf{v}}(\mathbf{k})\rangle = mN\mathbf{v} + \hbar\mathbf{k} \tag{11.153}$$

$$\langle \phi_{\mathbf{v}}(\mathbf{k})|H|\phi_{\mathbf{v}}(\mathbf{k})\rangle = E_0 + N\frac{mv^2}{2} + \omega_0(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v} \tag{11.154}$$

The result $\omega_0(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v}$ is quite reasonable: If $\omega_0(\mathbf{k})$ is the excitation spectrum with respect to the frame moving with the fluid, then the Doppler-shifted energy, as viewed by an observer at rest in the laboratory, is $\omega_0(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v}$.

Of course, the excitation spectrum should be $\omega(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v}$ rather than $\omega_0(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v}$. This improvement can be achieved by including the self-energy in (11.147) which produces the improved excitation spectra. This step is done and the excitations are indeed the ones which are physically realistic.

The momentum operator gives $mN\mathbf{v} + \hbar\mathbf{k}$ for a single excitation of energy $\omega(\mathbf{k}) + \mathbf{k}\cdot\mathbf{v}$. In the actual system, there will be a number of excitations whose distribution is given by the

thermal occupation factor n_B for boson excitations. The average value of the momentum operator is

$$\langle \Psi_{\mathbf{v}} \mathbf{P} \Psi_{\mathbf{v}} \rangle_{\text{av}} = Nm\mathbf{v} + \frac{N}{\rho} \int \frac{d^3k}{(2\pi)^3} \hbar \mathbf{k} n_B[\omega(\mathbf{k}) + \mathbf{k} \cdot \mathbf{v}] \tag{11.155}$$

The argument of the boson distribution factor $n_B(Z) = 1/[\exp(\beta Z) - 1]$ is the excitation energy $\omega(\mathbf{k}) + \mathbf{k} \cdot \mathbf{v}$ in the laboratory frame. The chemical potential is zero in the superfluid state, since it is a form of Bose–Einstein condensation. Note that this expression is not the average momentum in the ground state of flowing fluid, since that would be found from the average of $m \int d^3k/(2\pi)^3 \hbar \mathbf{k} n_{\mathbf{k}-m\mathbf{v}} = Nm\mathbf{v}$, where $n_{\mathbf{k}-m\mathbf{v}}$ is the momentum distribution of the particles in the ground state of the superfluid. The total momentum in the ground state of the flowing fluid is just $Nm\mathbf{v}$. Instead, (11.155) is the total momentum in the flowing fluid as contributed by the ground state plus the excitations. The excitations are produced by the thermal fluctuations, so that the total momentum will be temperature dependent.

For small velocities the argument of the distribution function can be expanded in powers of $\mathbf{k} \cdot \mathbf{v}$: $n_B(\omega(\mathbf{k}) + \mathbf{k} \cdot \mathbf{v}) \approx n_B(\omega) + \beta \mathbf{k} \cdot \mathbf{v} n_B'(\omega) + O(v^2)$. The first term with just $n_B[\omega(\mathbf{k})]$ gives a zero average of $\hbar \mathbf{k}$, and the important result is from the second term in the series expansion:

$$\langle \mathbf{P} \rangle_{\text{av}} = Nm\mathbf{v} + \frac{\beta N}{\rho} \int \frac{d^3k}{(2\pi)^3} \hbar \mathbf{k}(\mathbf{k} \cdot \mathbf{v}) n_B'[\omega(\mathbf{k})] \tag{11.156}$$

Since the spectrum n_B' is isotropic in $\mathbf{k}$, the angular integrals are done easily and give a vector in the direction $\mathbf{v}$ in $\int d\Omega_k \mathbf{k}(\mathbf{k} \cdot \mathbf{v}) = 4\pi \mathbf{v} k^2/3$. Collect together the remaining factors shows that the total momentum is expressed in terms of the normal fraction of superfluid which was defined in (11.148):

$$\langle \mathbf{P} \rangle_{\text{av}} = Nm\mathbf{v}\left(1 - \frac{\rho_N}{\rho}\right) = \frac{\rho_S}{\rho} Nm\mathbf{v} \tag{11.157}$$

The final result is that the total momentum in the flowing fluid is given only by the fraction ρ_S/ρ which is superfluid. The normal fraction ρ_N/ρ gives no contribution to the momentum.

In a normal liquid, uniform flow is not an eigenstate because viscosity damps the flow. The superfluid has no viscosity, and uniform flow is an eigenstate. At nonzero temperatures $0 < T < T_\lambda$, only the superfluid component ρ_S/ρ can flow without damping and is an eigenstate with a well defined momentum.

The physical picture of this result is very subtle. The flow of liquid is down a pipe, which remains in the laboratory frame. The normal component of the fluid is pinned by the pipe and does not participate in the flow. There are two important points to this derivation. The reference point for the excitation spectra is the rest frame of the superfluid. The excitations are made from this flowing fluid and have an energy $\omega(\mathbf{k})$ with respect to the rest frame of the superfluid. However, the influence of the excitations was calculated in the rest frame of the walls of the container, in this case the pipe. The coordinates were shifted to the laboratory frame and which gave the Doppler-shifted energy $\omega(\mathbf{k}) + \mathbf{k} \cdot \mathbf{v}$ before doing the thermal averaging. The rest frame of the container and the superfluid both enter the calculation. The result is that only the superfluid density carries momentum when the walls are stationary.

The flowing superfluid is in an eigenstate of the system, which can persist without damping. The state $\Psi_{\mathbf{v}}$ has an eigenvalues $(\rho_S/\rho)Nm\mathbf{v}$ of the momentum operator. In a completely normal fluid, such as neon or ^{4}He for $T > T_\lambda$, uniform flow down a fixed pipe is

not an eigenstate, and the flow is dissipated by viscosity. In the superfluid, the component ρ_S is not dissipated.

This calculation should be compared with other similar systems. What if the entire container of superfluid is just picked up and moved at a constant velocity; i.e., what if a Dewar is driven on a truck? Then the walls and the superfluid are moving at the same velocity, and the thermal averaging should be done in their mutual rest frame, which is moving with the truck. Then the momentum average gives

$$\langle \mathbf{P} \rangle_{\text{av}} = Nm\mathbf{v} + \frac{N}{\rho} \int \frac{d^3k}{(2\pi)^3} \hbar \mathbf{k} n_B[\omega(\mathbf{k})] = Nm\mathbf{v} \tag{11.158}$$

Now the result is $Nm\mathbf{v}$ since both normal and superfluid components are moving with the same velocity.

11.2. LIQUID ^{3}He

It is certainly a misnomer to call liquid ^{3}He a "normal liquid." The title is used here to distinguish it from the superfluid phase of the liquid. This system is interesting because its properties are so exceptional. It is really the only ideal Fermi liquid in earthbound nature, since electrons in metals are not ideal because their properties are perturbed by the crystal lattice of ions. The ^{3}He atoms are charge neutral, so that the properties of the Fermi system are much different from electrons in metals. They are also strongly interacting and highly correlated, so the noninteracting Fermi gas is not a good starting point for a description of its ground state. The discovery that ^{3}He has a superfluid phase, which is akin to the BCS superconductor, has heightened interest in this liquid.

The theory of these systems operates at several levels. One example is the two-fluid model of hydrodynamic flow, and another example is the microscopic derivation of $S(k)$ and $g(r)$. The theoretical description can be viewed as a sequence of plateaux, where each plateau is a self-consistent description of the system. For example, the excitation spectrum of liquid ^{4}He is such a plateau. One feature of these plateaux is that they can be compared directly with experiment. The theoretical effort generally is devoted to deducing the theory of one plateau by using the model of the lower one. For example, in liquid ^{4}He, one does not try to derive in one step the hydrodynamic flow properties from the microscopic theory. Instead, one uses microscopic theory to find the excitation spectrum $\omega(k)$, which is one plateau. Then one uses Landau theory to derive the two-fluid model in terms of $\rho_S(T)$ and $\rho_N(T)$, which are described in terms of $\omega(k)$. The theory is a sequence of plateaux, each of which can be compared with experiment, and one uses the theory to try to relate one step to the next one higher in the sequence.

The same sequence applies to the properties of liquid ^{3}He. The first, and most important plateau, is the Fermi liquid theory which was developed by Landau (1946, 1956, 1957), who has provided the main insight into both ^{4}He and ^{3}He. This theory was developed as a phenomenological model for the behavior of low-lying excitations of the strongly interacting Fermi system. These excitations can be described by a few parameters, which can be deduced from experiments. A major goal of microscopic theory is to derive these parameters from first principles. The major effort along this line, as was also the case for ^{4}He, is by Feenberg and his associates. So far they have achieved a good level of agreement, which is on the order of 30% at the worst and is better for some quantities. Since the Landau theory of the Fermi liquid

is the major plateau of the theory, this section is devoted to describing it as well as summarizing the microscopic theory.

The next plateau is a description of the superfluid state, which is usually done using the parameters of Fermi liquid theory. These parameters play a central role in the theoretical development. The superfluid properties are well described in this fashion using the experimentally measured parameters. In this sense a quantitative theory is available for the superfluid state.

11.2.1. Fermi Liquid Theory

There are excellent reviews of Fermi liquid theory as applied to liquid ^{3}He in the normal phase. Several of these are Pines and Nozieres (1966), Baym and Pethick (1978), Ron (1975), and Leggett (1975). These treatments are usually more detailed than the one here, and the reader is encouraged to read some of these others.

The symbol $n_{\mathbf{p}}$ is defined in (3.135) as the average number of atoms in the Fermi liquid with momentum $\mathbf{p}$. In a noninteracting Fermi system, at zero temperature, this distribution is $n_{\mathbf{p}} = \Theta(k_F - |\mathbf{p}|)$. Of course, in a strongly interacting Fermi system, this distribution will be strongly smeared. Figure 11.11 shows the momentum distribution for ^{3}He measured by neutron scattering at $T = 0.37$ K by Mook (1985). The points are experimental. The solid line is a fit to the Fermi distribution $n(p) = 1/[\exp(\xi_p/k_B T_K) + 1]$ with $T_K = 1.8$ K. The smearing caused by particle–particle interactions can be represented by an effective temperature in the Fermi distribution.

Throughout the remaining part of this chapter, the symbol $n_{\mathbf{p}}$ is used to mean something entirely different from its usage elsewhere in this book. Whereas before $n_{\mathbf{p}}$ always is the distribution of states with free-particle momentum $\mathbf{p}$, now it has a different meaning. Fermi liquid theory has a well-entrenched set of notation, with a universal set of symbols, whose meaning is well agreed upon. These strong conventions are followed here, and now $n_{\mathbf{p}}$ is the distribution of excitations of momentum $\mathbf{p}$.

The objective of Fermi liquid theory is to describe the low-lying excited states of the fermion system. There is a ground state which the system would be in at zero temperature except for superfluid phases and this ground state will be discussed in the next section. In analogy with the electron gas, two kinds of low-lying excitations are expected. The first are density oscillations, which are plasmons in the charged electron gas. In the neutral Fermi liquid they are sound waves. The second type of excitation is particle–hole pairs. There is also a third kind of excitation, which is a damped spin wave, called a paramagnon. Fermi liquid

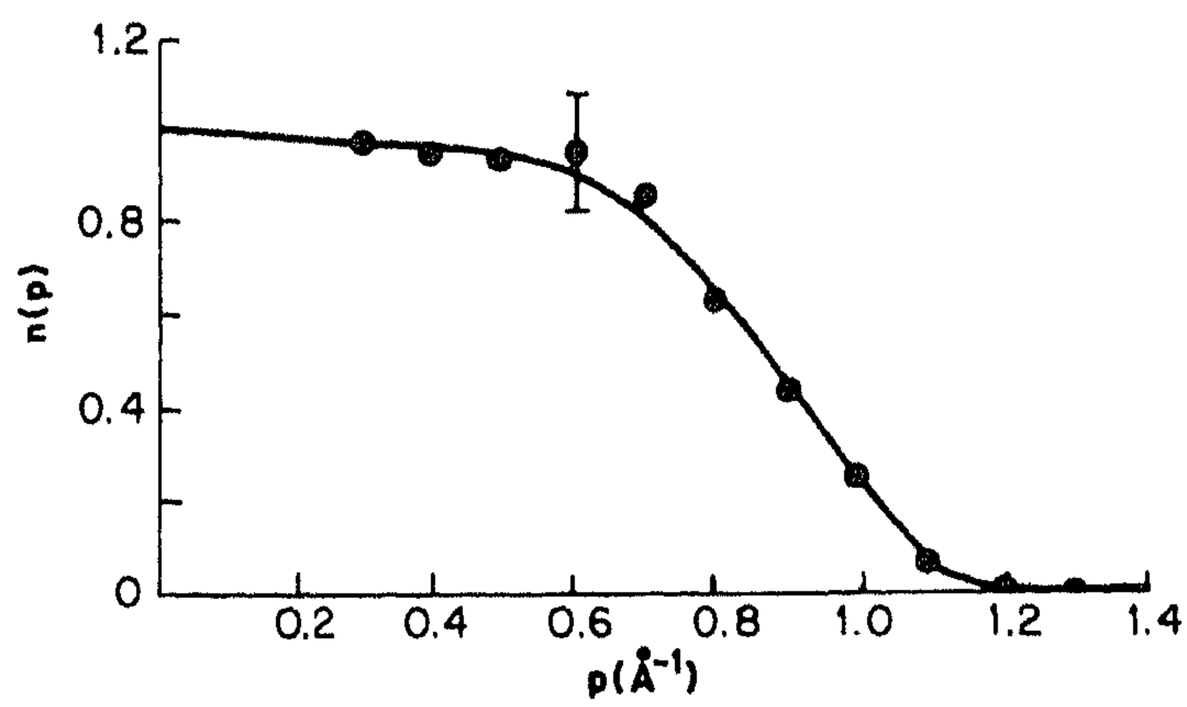


FIGURE 11.11 Momentum distribution n_p of particles in ^{3}He at $T = 0.37$ K. The solid line is a least-squares fit to the Fermi distribution with an effective temperature of $T_p = 1.8$ K. *Source:* Mook (1985) (used with permission).

theory is primarily aimed at a description of the properties of low-lying particles and holes. However, the other excitations—density oscillations and paramagnons—can be viewed as collective resonances of these particle and hole excitations. Fermi liquid theory also describes them. Indeed, many of the so-called Fermi liquid effects actually come from the collective properties. The discussion will follow the historical pathway and first treat the classical Fermi liquid theory due to Landau. Then the modern interpretation in terms of collective properties will be discussed.

By using the parameters in Table 11.1, one would calculate a Fermi energy of $E_{F0} = 4.97$ K for a noninteracting gas of the same density and mass of ^{3}He. The actual degeneracy energy in liquid ^{3}He is about two-thirds of this value. Fermi liquid theory is valid only for the particle–hole excitations whose energy is a small fraction of the degeneracy energy—say 5–10%. The excitations in this theory have low energy, say less than 0.1–0.2 K. If these excitations are used to describe the thermodynamics of the excited states of the system, it can only apply for temperatures below 0.1–0.2 K. Otherwise, the thermodynamic properties will involve states beyond the range of the theory. Fermi liquid theory is concerned with very low-energy excitations, which are important only in the liquid at very low temperatures. Since the superfluid properties begin at a temperature two decades lower than this, there is an appreciable temperature range over which Fermi liquid theory is applicable.

In the homogeneous electron gas, the imaginary part of the electron self-energy, from electron–electron interactions, vanishes for electrons whose energy is right at the chemical potential μ. Furthermore, Im Σ is also small for states near the chemical potential, since it vanishes as $(\varepsilon_{\mathbf{p}} - \mu)^2$. The low-lying excited states of the Fermi liquid have a long lifetime. The same phenomenon is assumed to happen in ^{3}He, and general theorems can be proved which demonstrate this with rigor (Luttinger and Nozieres, 1962). In cases where $\mathrm{Im}\,\Sigma \approx 0$, the spectral function $A(p, \omega)$ is sharply peaked and has the character of a delta function. There is a unique relationship between energy and momentum in these cases, and the excitation is satisfactorily described by only one of these two variables. In Fermi liquid theory this variable is conventionally taken to be the momentum $\mathbf{p}$. An excitation of momentum $\mathbf{p}$ is assigned an energy $\varepsilon_{\mathbf{p}}$.

The energy zero is defined where the excitation energy $\varepsilon_{\mathbf{p}}$ is zero at $\mathbf{p} = 0$. Furthermore, if the energy is expanded in a Taylor series about point $\mathbf{p} = 0$, the odd terms in this series p^{2m+1} must vanish because of the homogeneity of the liquid and the isotropy between $\mathbf{p}$ and $-\mathbf{p}$. The first term in the expansion is p^2, and other terms are usually ignored. Write $\varepsilon_{\mathbf{p}} \propto p^2$, and the constant of proportionality defines the effective mass m^*, $\varepsilon_{\mathbf{p}} = p^2/2m^*$. In liquid ^{3}He, the effective mass has a value of $m^* = 2.76m$, where m is the bare mass of ^{3}He. These and other experimental properties are reviewed by Glyde (1994) and by Vollhardt and Wölfe (1990).

TABLE 11.1 Properties of liquid helium (pressure = 1 bar). The parameters σ, ε are for the Lennard-Jones potential

Parameter	^{4}He	^{3}He
m	6.65×10^{-27} kg	5.01×10^{-27} kg
ρ	2.18×10^{28} atoms/m^3	1.64×10^{28} atoms/m^3
$C_1(T = 1\text{ K})$	238 m/s	182 m/s
σ	2.648 Å	2.648 Å
ε	1.484×10^{-22} J = 10.7 K	1.484×10^{-22} J = 10.7 K
E/N	−7.20 K/atom	−2.52 K/atom

The ground state of the fermion system does not allow all the excitations to be in the state with $\mathbf{p} = 0$. The exclusion principle forces many excitations into states of nonzero $\mathbf{p}$, and all energy states are occupied at zero temperature for $\varepsilon_{\mathbf{p}} < \mu$. Define $n_{\mathbf{p}}^0$ as the number of excitations of momentum $\mathbf{p}$ in the ground state, in which case the ground state energy is

$$E_g = E_0 + \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} n_{\mathbf{p}}^0 \tag{11.159}$$

where E_0 is the energy when all the quasiparticles are in $\mathbf{p} = 0$, which would be the case if they were bosons. Of course, the result (11.159) is actually nonsense, since the approximation of treating the excitations as free particles with an effective mass m^* and no damping is valid only when $|\varepsilon_{\mathbf{p}} - \mu| \leq k_B T$ and does not apply throughout the degenerate Fermi sea. The quantity $n_{\mathbf{p}}^0$ is a totally fictitious concept. However, it is never used for any property of $n_{\mathbf{p}}^0$ except at the chemical potential, where the following form is assumed at nonzero temperature:

$$n_{\mathbf{p}}^0 = n_F(\varepsilon_{\mathbf{p}} - \mu) = \frac{1}{e^{\beta(\varepsilon_{\mathbf{p}} - \mu)} + 1} \tag{11.160}$$

This approximation is surely valid only very near the chemical potential $\beta(\varepsilon_{\mathbf{p}} - \mu) \leq 1$, which is the only place it is used.

The important feature is the excitation of the system above the ground state. When excitations are present, the distribution $n_{\mathbf{p}}$ of these excitations will be different from the ground state distribution $n_{\mathbf{p}}^0$. The difference between these two distributions is defined as $\delta n_{\mathbf{p}} = n_{\mathbf{p}} - n_{\mathbf{p}}^0$. The important physical idea is that the crucial quantity is neither $n_{\mathbf{p}}$ nor $n_{\mathbf{p}}^0$ but rather $\delta n_{\mathbf{p}}$, since it gives the number of excitations in the excited state which is important for the thermodynamics of the system at low temperature. Neither $n_{\mathbf{p}}$ nor $n_{\mathbf{p}}^0$ can be determined with great accuracy, but $\delta n_{\mathbf{p}}$ can be determined, which is all that matters. For example, the total energy of the system, including the excited states, is

$$E = E_0 + \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} n_{\mathbf{p}} \tag{11.161}$$

$$= E_0 + \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} n_{\mathbf{p}}^0 + \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} [n_{\mathbf{p}} - n_{\mathbf{p}}^0] \tag{11.162}$$

$$= E_g + \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} \delta n_{\mathbf{p}} \tag{11.163}$$

The excited state energy is $\sum \varepsilon_{\mathbf{p}} \delta n_{\mathbf{p}}$.

The important quantity is not the energy but the free energy $F = E - \mu N$, where N is the total number of particles. The particle number is defined in terms of the number of particles in the ground state N_0

$$N_0 = \sum_{\mathbf{p}\sigma} n_{\mathbf{p}}^0 \tag{11.164}$$

$$N = \sum_{\mathbf{p}\sigma} n_{\mathbf{p}} = \sum_{\mathbf{p}\sigma} n_{\mathbf{p}}^0 + \sum_{\mathbf{p}\sigma} \delta n_{\mathbf{p}} = N_0 + \delta N \tag{11.165}$$

$$F = E - \mu N = E_g - \mu N_0 + \sum_{\mathbf{p}\sigma} (\varepsilon_{\mathbf{p}} - \mu) \delta n_{\mathbf{p}} \tag{11.166}$$

There are two key assumptions contained in these equations. The first is that the chemical potential μ does not change because of the excited states, although it does change with temperature. The second assumption is that there is a one-to-one correspondence between quasiparticles and the excited states of the noninteracting Fermi gas. Adding a quasiparticle also adds a particle and changes N by unity. Of course, usually there are about equal numbers

of quasiparticles above $\varepsilon_{\mathbf{p}} > \mu$, which have $\delta n_{\mathbf{p}} > 0$, as there are a depletion of them for $\varepsilon_{\mathbf{p}} < \mu$, which have $\delta n_{\mathbf{p}} < 0$. The latter excitations are called holes.

So far the Fermi liquid theory has been equivalent to calling the excitations free particles with an effective mass m^*. If that were all of it, the theory would be unexceptional and not quantitatively accurate. The genius of Landau was in realizing that the expansion for the free energy must have another term, which is due to the interactions. This contribution is required for a self-consistent theory. The formulas derived so far are the terms in an expansion of the free energy in the excited distribution $\delta n_{\mathbf{p}}$. That is, $F(\delta n_{\mathbf{p}})$ is a functional of $\delta n_{\mathbf{p}}$ and can be expanded formally in a Taylor series in this parameter. The Landau theory assumes that $\delta n_{\mathbf{p}}$ is small and that an accurate quantitative theory is achieved after a few terms in this expansion.

The terms derived so far are given in (11.165). The first term is $F_0 = E_g - \mu N$. The next term appears to be first order in $\delta n_{\mathbf{p}}$. A closer inspection shows that this term is actually of order $(\delta n_{\mathbf{p}})^2$. The summation over $\mathbf{p}$ will run over all values for which $\delta n_{\mathbf{p}}$ is nonzero. However, the integrand contains $(\varepsilon_{\mathbf{p}} - \mu)$, which vanishes at $\varepsilon_{\mathbf{p}} = \mu$. If $\delta n_{\mathbf{p}}$ has value up to Δ away from the chemical potential, the kinetic energy integral $\Sigma_{\mathbf{p}}(\varepsilon_{\mathbf{p}} - \mu)\delta n_{\mathbf{p}} \propto \Delta^2$. This term is second order in the quantity $\delta n_{\mathbf{p}}$. Landau saw the need to include all terms of order $(\delta n_{\mathbf{p}})^2$. There is another term which describes the interactions between the excited quasiparticles and so contains the dependence $\delta n_{\mathbf{p}} \delta n_{\mathbf{p}'}$. This interaction is very dependent on the spin of the quasiparticles. The free-energy expansion has the form

$$F = F_0 + \sum_{\mathbf{p}\sigma}(\varepsilon_{\mathbf{p}} - \mu)\delta n_{\mathbf{p}\sigma} + \frac{1}{2\nu}\sum_{\mathbf{p}\mathbf{p}',\sigma\sigma'} f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}\delta n_{\mathbf{p}\sigma}\delta n_{\mathbf{p}'\sigma'} + O(\delta n)^3$$

where spin has been added to various quantities. This dependence on the particle spin is not due to the spin-dependent interactions. Indeed, the dipole–dipole forces between the nuclear moments are extremely small and may be neglected when discussing most phenomena. The spin dependence of the interaction is merely due to particle statistics. The antisymmetrization of the many-particle wave function for fermions causes an exchange hole around each particle. The exchange hole causes an effective exchange energy, just as it did for the electron gas. The exchange energy may be qualitatively understood by a consideration of the mutual scattering of two ^{3}He atoms. Since the two-particle wave function must be antisymmetric, the scattering properties depend on the total spin state of the two spin one-half particles as described in Sec. 11.1.1. If they are in an $S = 0$ state, the orbital cross section contains only even angular momentum components; if they have a total $S = 1$ state, only odd angular momentum components enter the cross section. The scattering and interaction depend on the relative spins of the two particles. It is a very important effect in liquid ^{3}He, since the effective interaction $f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$ is very spin dependent.

The spin dependence of the effective interaction is written as

$$f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'} = f^s_{\mathbf{p}\mathbf{p}'} + \boldsymbol{\sigma}\cdot\boldsymbol{\sigma}' f^a_{\mathbf{p}\mathbf{p}'} \tag{11.167}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ refer to Pauli spin matrices. Equation (11.167) is a very particular choice for the form of the interaction. Other possible spin dependencies are ignored, such as spin–orbit effects $(\boldsymbol{\sigma}\cdot\mathbf{p})(\boldsymbol{\sigma}'\cdot\mathbf{p}')$ or spin-other-orbit effects $(\boldsymbol{\sigma}\cdot\mathbf{p}')(\boldsymbol{\sigma}'\cdot\mathbf{p})$. The choice of a simple $\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'$ term derives from our understanding that the basic spin mechanism is due to exchange forces, which can be written this way.

Fermi liquid theory is really only accurate for describing the interactions of quasiparticles on the Fermi surface. In the interaction terms $f^s_{\mathbf{p}\mathbf{p}'}$ and $f^a_{\mathbf{p}\mathbf{p}'}$, the momentum variables have the magnitude k_F, and the only important variable is the angle θ between $\mathbf{p}$ and $\mathbf{p}'$. The

interaction terms are represented by an expansion in Legendre polynomials. It is also conventional to normalize the coefficients in this expansion by removing a factor of the density of states at the Fermi energy $N_F = m^* k_F/\pi^2\hbar^3$:

$$f^{s,a}_{\mathbf{pp}'} = \frac{1}{N_F}\sum_{l=0}^{\infty} F_l^{s,a} P_l(\cos\theta) \tag{11.168}$$

The coefficients F_l^s and F_l^a are sometimes called F_l and Z_l. They are the fundamental parameters in the theory and are dimensionless. They can be deduced from experiments, as shown later. Reliable experimental numbers are now available for F_0^s, F_1^s, F_0^a and F_1^a. A major goal of the microscopic theory is to derive them.

The term $f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$ describes the interaction between excited quasiparticles. The interactions cause the energy of an excited quasiparticle to depend on the number $\delta n_{\mathbf{p}\sigma}$ of other excited ones. The quasiparticle energy can be formally derived as a functional derivative of the total energy E:

$$E = E_g + \sum_{\mathbf{p}\sigma} \varepsilon_{\mathbf{p}} \delta n_{\mathbf{p}\sigma} + \frac{1}{2\nu}\sum_{\mathbf{pp}',\sigma\sigma'} f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'} \delta n_{\mathbf{p}\sigma}\delta n_{\mathbf{p}'\sigma'} \tag{11.169}$$

$$\frac{\delta E}{\delta(\delta n_{\mathbf{p}\sigma})} = \bar{\varepsilon}_{\mathbf{p}} = \varepsilon_{\mathbf{p}} + \frac{1}{\nu}\sum_{\mathbf{p}'\sigma'} f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}\delta n_{\mathbf{p}'\sigma'} \tag{11.170}$$

The energy of the excited quasiparticle is called $\bar{\varepsilon}_{\mathbf{p}}$. It depends on the number of other excited quasiparticles through the interaction term. One result of the interactions is that the total excitation energy is no longer equal to $\sum \bar{\varepsilon}_{\mathbf{p}}\delta n_{\mathbf{p}\sigma}$, since this expression overcounts the pairwise interaction term. The chemical potential μ is not altered by the density of excited quasiparticles as long as this density is very small compared to the density of atoms.

It is time to reconsider the distribution of excited quasiparticles. The same notation is used: i.e., that $n_{\mathbf{p}}$ is the total number of excited quasiparticles and that $n_{\mathbf{p}}^0$ is the number of excited quasiparticles in the ground state with energy $\varepsilon_{\mathbf{p}}$. Another symbol is needed to denote the quasiparticle density in terms of the actual energy variable $\bar{\varepsilon}_{\mathbf{p}}$. Call $\bar{n}_{\mathbf{p}}^0$ the equilibrium density of excited quasiparticles with energy $\bar{\varepsilon}_{\mathbf{p}}$, which is defined in the usual way in terms of the Fermi distribution function:

$$\bar{n}_{\mathbf{p}}^0 = n_F(\bar{\varepsilon}_{\mathbf{p}} - \mu) \tag{11.171}$$

$$\delta\bar{n}_{\mathbf{p}\sigma} = n_{\mathbf{p}\sigma} - \bar{n}_{\mathbf{p}}^0 \tag{11.172}$$

Similarly, the quantity $\delta\bar{n}_{\mathbf{p}\sigma}$ is the departure of the total number $n_{\mathbf{p}\sigma}$ from this equilibrium concentration. The two quantities $\delta n_{\mathbf{p}\sigma}$ and $\delta\bar{n}_{\mathbf{p}\sigma}$ are not independent and are related through the Fermi liquid parameters. This dependence can be derived by assuming that $\delta n_{\mathbf{p}\sigma}$ is a small quantity and expanding $\delta\bar{n}_{\mathbf{p}\sigma}$ in this parameter. The key step is the recognition that the difference $(\bar{\varepsilon}_{\mathbf{p}} - \varepsilon_{\mathbf{p}})$ is proportional to $\delta n_{\mathbf{p}\sigma}$

$$\delta\bar{n}_{\mathbf{p}\sigma} = n_{\mathbf{p}\sigma} - n_F(\bar{\varepsilon}_{\mathbf{p}} - \mu) \tag{11.173}$$

$$= n_{\mathbf{p}\sigma} - n_F(\varepsilon_{\mathbf{p}} - \mu + \bar{\varepsilon}_{\mathbf{p}} - \varepsilon_{\mathbf{p}}) \tag{11.174}$$

$$= n_{\mathbf{p}\sigma} - n_F(\varepsilon_{\mathbf{p}} - \mu) - (\bar{\varepsilon}_{\mathbf{p}} - \varepsilon_{\mathbf{p}})\frac{dn_F(\varepsilon_{\mathbf{p}} - \mu)}{d\varepsilon_{\mathbf{p}}} \tag{11.175}$$

$$\delta\bar{n}_{\mathbf{p}\sigma} = \delta n_{\mathbf{p}\sigma} - \frac{dn_F}{d\varepsilon}\frac{1}{\nu}\sum_{\mathbf{p}'\sigma'} f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}\delta n_{\mathbf{p}'\sigma'} \tag{11.176}$$

The first two terms in (11.175) are just $\delta n_{\mathbf{p}\sigma}$. The third term contains the factor $(\bar{\varepsilon}_{\mathbf{p}} - \varepsilon_{\mathbf{p}})$ which is also proportional to $\delta\bar{n}_{\mathbf{p}\sigma}$. It has the factor $-dn_F(\varepsilon_{\mathbf{p}} - \mu)/d\varepsilon_{\mathbf{p}}$, which becomes the delta function $\delta(\varepsilon_{\mathbf{p}} - \mu)$ at zero temperature. One can show that every term in Eq. (11.176) has this factor. Divide it out of the expression. Furthermore, the spin dependence of $\delta\bar{n}_{\mathbf{p}\sigma}$ is written as a symmetric or antisymmetric combination $\delta\bar{n}_{\mathbf{p}\pm} = \delta n^s_{\mathbf{p}} \pm \delta n^a_{\mathbf{p}}$. All these results are collected, and the distribution functions are expanded in spherical harmonic functions Y_{lm}. The first two equations below define $\delta n^{s,a}_{lm}$, etc., while the third equation is the integral equation (11.176):

$$\delta n_{\mathbf{p}\pm} = -\frac{dn_F(\varepsilon_{\mathbf{p}} - \mu)}{d\varepsilon_{\mathbf{p}}} \sum_{lm} Y_{lm}(\theta, \phi)(\delta n^s_{lm} \pm \delta n^a_{lm}) \tag{11.177}$$

$$\delta\bar{n}_{\mathbf{p}\pm} = -\frac{dn_F(\varepsilon_{\mathbf{p}} - \mu)}{d\varepsilon_{\mathbf{p}}} \sum_{lm} Y_{lm}(\theta, \phi)(\delta\bar{n}^s_{lm} \pm \delta\bar{n}^a_{lm}) \tag{11.178}$$

$$\delta\bar{n}^{s,a}_{lm} = \delta n^{s,a}_{lm} + 2\int \frac{d^3p'}{(2\pi)^3} \sum_{l'm'} \frac{d\Omega_p}{4\pi} \left(-\frac{dn_F}{d\varepsilon_{p'}}\right) Y^*_{lm}(\theta, \phi) f^{s,a}_{\mathbf{p}'\mathbf{p}} Y_{l'm'}(\theta', \phi')\delta n^{s,a}_{l'm'} \tag{11.179}$$

Integrate over angle and collect all terms with the same angular dependence, which provides the relationship between the two distributions of excited quasiparticles:

$$\delta\bar{n}^{s,a}_{lm} = \delta n^{s,a}_{lm}\left(1 + \frac{F^{s,a}_l}{2l+1}\right) \tag{11.180}$$

These results are useful later.

The distribution $\delta\bar{n}_{\mathbf{p}\sigma}$ seems more important than $\delta n_{\mathbf{p}\sigma}$. For example, Pines and Nozieres show that the particle current operator is given correctly by both of the following expressions:

$$\mathbf{J} = \sum_{\mathbf{p}\sigma} \frac{\mathbf{p}}{m^*} \delta\bar{n}_{\mathbf{p}\sigma} \tag{11.181}$$

$$\mathbf{J} = \sum_{\mathbf{p}\sigma} \frac{\mathbf{p}}{m} \delta n_{\mathbf{p}\sigma} \tag{11.182}$$

The two current operators (11.181) and (11.182) can be used to derive a relationship between m^* and m. The current operators are nonzero if the distribution functions $\delta\bar{n}_{\mathbf{p}\sigma}$ and $\delta n_{\mathbf{p}\sigma}$ have a nonzero component of angular momentum $l = 1$, say the value $l = 1, m = 0$. Then the two current expressions can be evaluated by using the definitions of $\delta\bar{n}_{\mathbf{p}\sigma}$ and $\delta n_{\mathbf{p}\sigma}$ in (11.180):

$$\frac{m^*}{m} = \frac{\delta\bar{n}^s_{10}}{\delta n^s_{10}} = 1 + \tfrac{1}{3}F^s_1 \tag{11.183}$$

A measurement of m^* gives the value of F^s_1. It was previously remarked that $m^* \approx 2.76m$, so that $F^s_1 \approx 5.3$. The Fermi liquid parameters can be large relative to unity.

Table 11.2 shows some predictions of Fermi liquid theory regarding the effect of the quasiparticle interactions upon measurable quantities. The first is the specific heat, which is enhanced by the effective mass m^*, as it is for the electron gas. The quantity F^s_1 is obtained by measuring this quantity. The isothermal compressibility and also the velocity of ordinary sound are changed by the ratio of $(1 + F^s_0)/(1 + F^s_1/3)$. For a noninteracting gas of ^{3}He particles, one would estimate the sound velocity from the formula $c_0^2 = v_F^2/3$ to be 95 m/s. The experimental value is 182 m/s. The ratio is $c/c_0 = 1.91$ gives the value of $F^s_0 = 9.15$ since F^s_1 is known. A recent compilation of these parameters is found in Table 11.3, as given

TABLE 11.2 Fermi liquid theory: $N_{F0} = mk_F\pi^2\hbar^3$

Quantity	Free particles	Fermi liquid theory
Specific heat	$C_{V0} = \frac{\pi^2}{3}k_B^2 T N_{F0}$	$\frac{C_V}{C_{V0}} = \frac{m^*}{m} = 1 + \frac{1}{3}F_1^s$
Compressibility	$\frac{1}{K_f} = \frac{n_0^2}{N_{F0}}$	$\frac{K}{K_f} = \frac{1+F_0^s}{1+F_1^s/3}$
Sound velocity	$c_0^2 = \frac{n_0}{mN_{F0}} = \frac{v_F^2}{3}$	$\left(\frac{c}{c_0}\right)^2 = \frac{K}{K_f}$
Spin susceptibililty	$\chi_0 = \mu_0^2 N_{F0}$	$\frac{\chi}{\chi_0} = \frac{1+F_1^a/3}{1+F_0^a}$

TABLE 11.3 Fermi liquid parameters: $A_\ell = F_\ell/[1 + F_\ell(2\ell + 1)]$ (from Greywall 1983)

Parameter	Experiment	Parameter	Experiment
F_0^s	9.15	A_0^s	0.90
F_1^s	5.27	A_1^s	1.91
F_0^a	−0.70	A_0^a	−2.33
F_1^a	−0.55	A_1^a	−0.67

by Greywall (1983) who found that $m^* = 2.76m$. The other parameters $A_l^{s,a}$ shown in Table 11.3 are explained later.

The other interesting number is F_0^a, which is the isotropic term in the spin-dependent part of the interaction. It is deduced from the spin susceptibility, which is found to be much larger than the free-particle value χ_0. The enhancement of χ over χ_0 suggests that F_0^a is negative and near unity, so that the denominator is small in the theoretical expression for the susceptibility. Later it is shown that this has an important consequence for the theory of superfluidity. The large positive value of F_0^s and the negative value of F_0^a combine to suppress the quasiparticle interaction in the s-wave channel and enhance it in the p-wave channel. It is this reason that the superfluid phases in ^{3}He pair up in the spin triplet state, which has a p-wave orbital part.

11.2.2. Experiments and Microscopic Theories

The microscopic theories of liquid ^{3}He are even more complicated than the theories of ^{4}He, because the ground state is more complicated. The complexity is due to the anti-symmetrization of the many-particle wave function. If this ground state wave function is defined as $\Psi_g(x_1, \ldots, x_N)$, where $x_j = (\mathbf{r}_j, \sigma_j)$ are general coordinates for both position and spin, then the wave function changes sign under the exchange of any pair of coordinates x_i and x_j:

$$\Psi_g(x_1, x_2, \ldots, x_i, \ldots, x_j, \ldots, x_N) = -\Psi_g(x_1, x_2, \ldots, x_j, \ldots, x_i, \ldots, x_N) \quad (11.184)$$

For a noninteracting system of fermions, the particle states are described by single-particle orbitals $\phi_p(x_j)$, and the many-particle wave function in the Hartree–Fock approximation is a

Slater determinant of these orbitals. The determinant has the property that no two particles can occupy the same single-particle orbital state, or else the wave function is trivially zero.

The nature of the ground state wave function is less clear in this strongly interacting and highly correlated liquid. The microscopic picture of the liquid has the atoms bouncing around inside of a small cage defined by the average position of neighboring atoms. This picture is correct for liquid ^{4}He and leads to a correlated wave function of the CBF form, which is highly successful for describing bosons. It is not nearly so obvious how to construct a ground state wave function for highly correlated fermions. There probably should be something equivalent to the CBFs but for fermions. The question is how the fermion nature of the particles is expressed in real space. When visualizing the atoms bouncing around, how is this motion different when they are fermions rather than bosons? How does one construct a CBF which has the fundamental antisymmetrization property (11.184)?

Feenberg and his associates have provided the most successful microscopic theory to date, which compares favorably to the available experiments. Their ground state wave function for liquid ^{3}He has the form of an algebraic product of two familiar wave functions: the CBF with the form used for ground state of ^{4}He and a Slater determinant of plane-wave states with spin:

$$\Psi_g(x_1, x_2, \ldots, x_N) = \Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)\Phi_{p_1\cdots p_N}(x_1, \ldots, x_N)$$
$$\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = L_N \exp\left[-\sum_{i>j}\left(\frac{b}{|\mathbf{r}_i - \mathbf{r}_j|}\right)^5\right] \tag{11.185}$$

$$\Phi_{p_1\cdots p_N}(x_1, \ldots, x_N) = \det|\phi_{p_1}(x_1)\phi_{p_2}(x_2)\cdots| \tag{11.186}$$

The CBF was taken to be the form (11.62) of a product of pairwise correlations, with a single short-range repulsive term as in (11.77). The value of b changes slightly from the value used in ^{4}He, because of the mass difference. The Slater determinant is the same one which would be used for a system of noninteracting particles and contains single-particle wave functions $\phi_p(x)$ which are plane-wave states and spin functions for spin up (α) or down (β). The Slater determinant provides the necessary antisymmetrization, while the CBF provides the necessary short-range correlation in the atomic motion.

This wave function is used to calculate ground state properties such as the energy $E_g = \langle g|H|g\rangle/\langle g|g\rangle$, $S(k)$, etc. These calculations are extremely complicated and will not be described here. Interested readers are referred to Feenberg's book (1969). All many-particle matrix elements are evaluated by a cumulant expansion which must be carried to high order. They involve the evaluation of three- and four-particle correlations in the liquid state. Some of their results will now be summarized.

The ground state energy is discussed first. Feenberg and his associates first calculated the ground state energy assuming ^{3}He was a boson. That is, they evaluated $E_0 = \langle\Psi_0|H|\Psi_0\rangle$ where Ψ_0 is the CBF for liquid ^{3}He. Since this wave function is symmetric under coordinate exchange, E_0 is the energy of the equivalent boson liquid. The value of E_0 differs somewhat for the different ways of doing the cumulant expansion and estimating three- and four-particle correlations but has the typical value of $E_0 = -2.9$ K/atom. A boson liquid with the mass of ^{3}He is less bound than liquid ^{4}He, which reflects the increased zero-point motion of the lighter mass. Although the density of liquid ^{3}He is lower than liquid ^{4}He, the average value of kinetic energy per particle is nearly the same value of 13.4 K/atom. The average value of potential energy is -16.3 K/atom, giving the difference of -2.9 K/atom. Thus the average

kinetic energy per atom is still high, and the particles still have short-range fluctuations in their position within the cage defined by the neighboring atoms.

The next calculation is to add the Slater determinant of the single-particle orbitals to the ground state wave function and to compute the new ground state energy. It is defined as $E_g = E_0 + \delta E$, where δE is the change in energy due to the motion of the particles in the Fermi sea. For a gas of noninteracting particles, this change would be $\delta E = 3E_{F0}/5 = 0.6(4.97) = 3.0$ K. Feenberg and his associates find the smaller value of $\delta E = 1.7$ K/atom, which is a small number compared to the average kinetic energy in the boson ground state, $\approx$14 K/atom. The short-range motions of the particles are more important than the kinetic energy of motion in the Fermi sea. The final prediction of the ground state energy per atom $E_g = -1.2$ K/atom is only in fair agreement with the experimental value of -2.52 K/atom. However, the least accurate theoretical number is probably E_0. There could easily be a 1 K error in the calculation of this quantity, as there was in the similar calculation for ^{4}He, where it was less noticeable as the difference between 6 and 7 K/atom. This number is difficult to find accurately because of the large cancellation between the kinetic and potential energies and their sensitivity on the trial function for the pairwise part of the CBF. The agreement between theory and experiment, in liquid ^{3}He, must be regarded as satisfactory.

The next ground state property is the liquid structure factor $S(k)$. It can be measured by both X-ray and neutron scattering. Figure 11.12 summarizes theory and X-ray data. The points are experimental, and the lines are the theories of Massey (1966) and Massey and Woo (1967). The agreement is obviously excellent. Since the theories were published as predictions several years before the experiments, the theoretical success is even greater. There is great similarity between $S(k)$ for liquid ^{3}He and ^{4}He. Both have peaks near $k \sim 2.0\,\text{Å}^{-1}$ and then rapidly approach unity with small oscillations at large values of k. The $S(k)$ values for ^{3}He are about 40% higher than for ^{4}He in the low k region. The good agreement between theory and experiment for $S(k)$ means that similar good agreement exists between their Fourier transforms, which are the pair distribution functions $g(r)$.

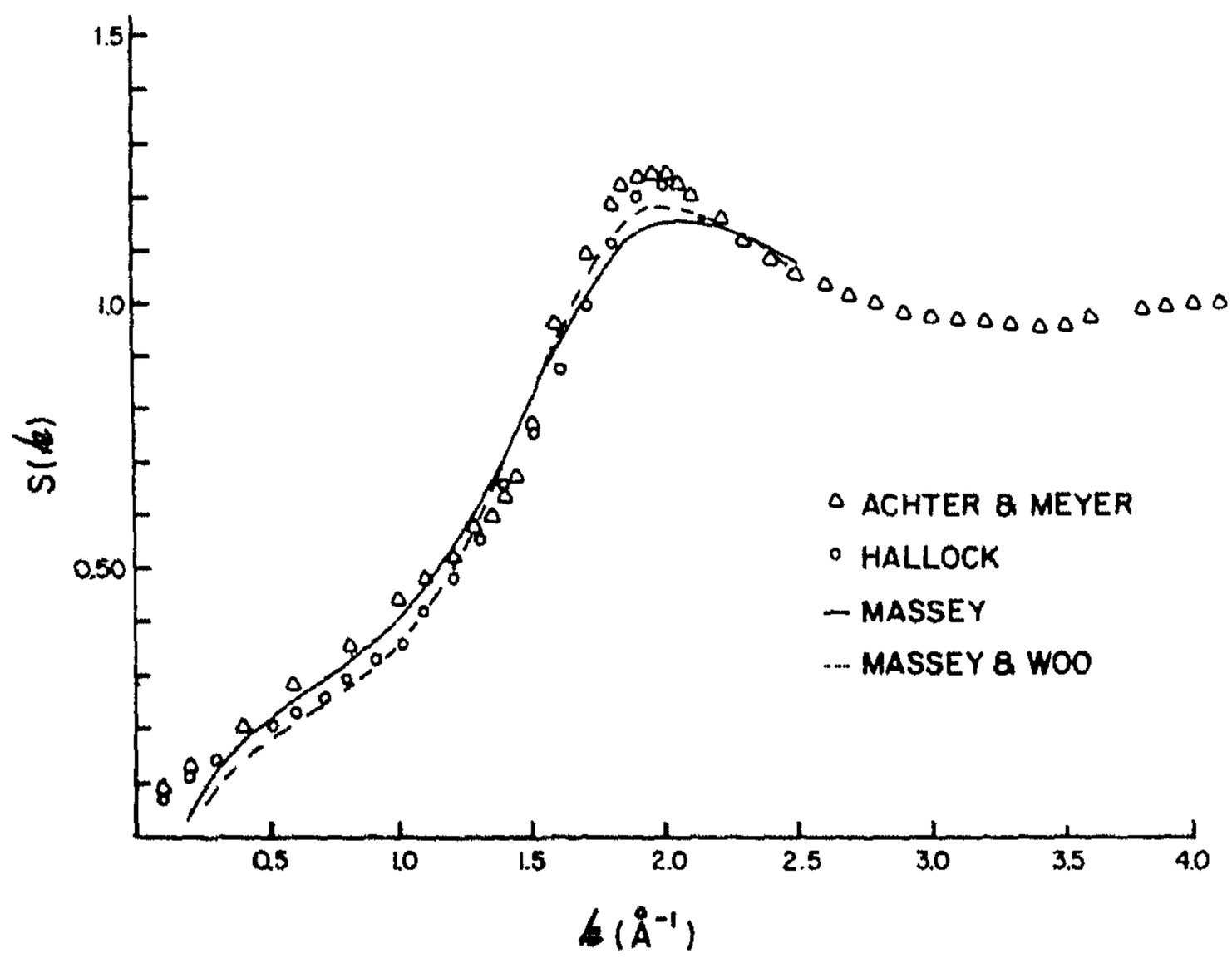


FIGURE 11.12 Plot of $S(k)$ for liquid ^{3}He. The points are the X-ray scattering data of Achter and Meyer (1969) and Hallock (1972). The solid curves are the theories of Massey (1966) and Massey and Woo (1967). The agreement is obviously excellent.

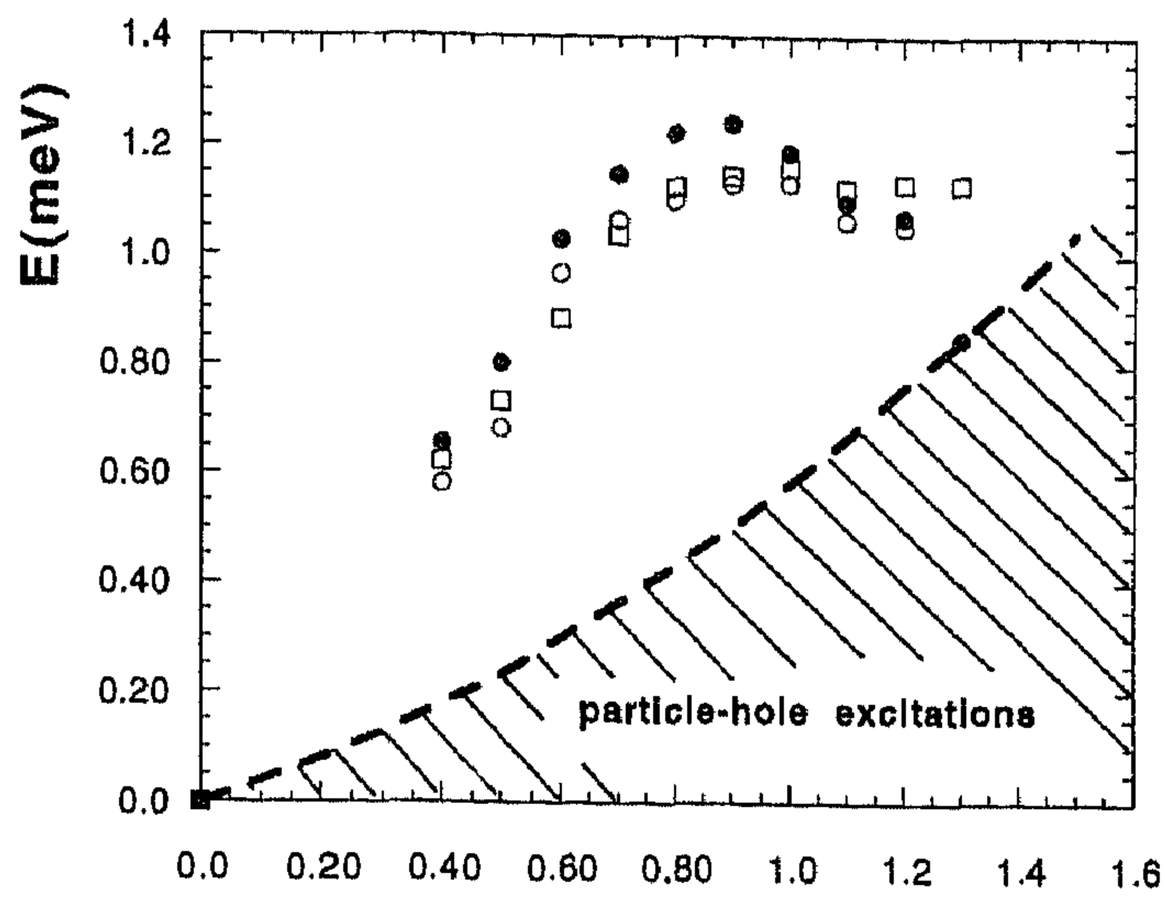


FIGURE 11.13 Zero sound mode energy in liquid ^{3}He. The points are from Sköld and Pelizzari (1980). From Glyde (1994, used with permission).

The liquid must have a long-wavelength collective oscillation. In ordinary liquids it is just the longitudinal sound wave. In the electron gas it is at the plasma frequency ω_p, where the charge of the electron gas makes the mode have a nonzero frequency even in the limit where $q \to 0$. Since liquid ^{3}He is electrically neutral, the long-wavelength density oscillation will just be a sound wave, which is called the *zero sound mode*. Figure 11.13 shows the zero sound mode in liquid ^{3}He as measured by neutron scattering. The cross-hatched region in the lower corner of the figure are the particle–hole pair excitations out of the Fermi sea, as is found in metals. They are calculated using $m^* = 3m$.

There are several different kinds of sound waves which can propagate in these systems. The ordinary sound waves at very long wavelengths are called first sound. Their speed is calculated by including the particle collisions, since such collisions are very frequent compared to the sound frequency. This results in a hydrodynamic version of the sound propagation. At much higher sound frequencies, the collision rate between particles becomes small compared to the sound frequency, and the speed must be calculated in the collisionless approximation. This regime is called zero sound and has a higher velocity.

The third excitation of liquid ^{3}He are the spin fluctuations at low energy. These will be treated fully in the following section.

11.2.3. Interaction Between Quasiparticles: Excitations

Fermi liquid theory is not a complete description of all the dynamical motion of quasiparticles in liquid ^{3}He. The theory was constructed by Landau to explain thermodynamic and static measurements such as specific heat and spin susceptibility, and it does that very well. In order to explain other measurements, it must be decided whether enough information is available in the Fermi liquid parameters or whether additional facts are needed about the quasiparticles. The additional measurements are transport experiments such as thermal conductivity and viscosity. A theory is also needed for low-energy spin fluctuations and also superfluidity. So far all that is known about the quasiparticles are their effective mass and the four interaction parameters $F_{0,1}^{s,a}$. It must be decided what these parameters mean in terms of the microscopic scattering theory, and what additional information is needed to fully describe the interaction between quasiparticles.

It is useful to remember the main features of electrons in metals, which is the most understood Fermi liquid. The electrons at the Fermi surface have an effective mass m^* which is comprised of several contributions. Those of the electron–electron interactions, and also the band structure, gave a contribution to m^*/m which was not very energy dependent. However, the contribution from electron–phonon interactions was often very large and varied strongly with energy, so that it was negligible a Debye energy away from the Fermi surface. In liquid ^{3}He, one should ask the question whether the ratio $m^*/m \approx 3$ changes rapidly or slowly as the energy of the quasiparticles is varied. Or, to rephrase the question, what is the appropriate energy scale which determines the variation of these parameters? Another motivation for this question is provided by recalling the BCS equation for the transition temperature of a superconductor:

$$k_B T_c = 1.14\hbar\omega_D \exp\left(-\frac{1}{N_F V_0}\right) \tag{11.187}$$

The exponent has the factors $N_F V_0$ which are the density of states and the effective interaction between quasiparticles at the Fermi surface. These two parameters can be estimated from the parameters of Fermi liquid theory. However, the prefactor in this equation contains the Debye energy $\hbar\omega_D$, which is the effective energy range over which the phonon attraction between quasiparticles is operative. This effective energy range of the interaction cannot be obtained from the parameters of Fermi liquid theory and must be deduced from other considerations.

In elemental metals, superconductivity is usually caused by the electron–phonon interaction. This theory has been made very quantitative. The accuracy is helped by a number of smallness parameters, such as c_s/v_F, ω_D/E_F, and m/M, in which the theoretical expressions can be expanded. None of these apply to liquid ^{3}He. It has $c_s > v_F$, and the energy of high frequency sound exceeds the Fermi energy of $E_F = k_B T_F \approx 3$ K = 0.26 meV. In fact, no obvious smallness parameter comes to mind. Nevertheless, there probably is one since the ratio of $T_c/T_F \approx 10^{-3}$ is similar between metals and liquid ^{3}He.

The scattering of two quasiparticles can be described in the following manner. Two quasiparticles initially have momenta $\mathbf{p}_1$ and $\mathbf{p}_2$. They interact and exchange momentum $\mathbf{q}$ and end up in the states $\mathbf{p}_3 = \mathbf{p}_1 + \mathbf{q}$ and $\mathbf{p}_4 = \mathbf{p}_2 - \mathbf{q}$. The scattering of the quasiparticles is described by a phenomenological matrix element $M_{\sigma\sigma'}(\mathbf{p}_1, \mathbf{p}_2; \mathbf{p}_3, \mathbf{p}_4)$. The spin indices are included, since the effective scattering is spin dependent. The goal is to obtain a quantitative description of this matrix element. It is defined to include exchange events, so it describes the direct scattering plus the exchange scattering.

The first step is to find the relationship between this matrix element and the Fermi liquid parameters. This identification is found by an intuitive argument. In a weakly interacting Fermi gas, the Hartree energy is given by

$$\Sigma_H = \frac{1}{\nu}\sum_{\mathbf{p}'} n_F(\xi_{\mathbf{p}'}) V(\mathbf{q} = 0) \tag{11.188}$$

The self-energy diagram for Σ_H has a particle of momentum $\mathbf{p}$ scatter from the other particles with momentum $\mathbf{p}'$ and density $n_F(\xi_{\mathbf{p}'})$. No momentum is exchanged, so the self-energy is expressed in terms of the potential $V(0)$ at zero-momentum transfer ($\mathbf{q} = 0$), which is equivalent to the forward scattering amplitude for the two particles $\mathbf{p}$ and $\mathbf{p}'$. The exchange

self-energy $\Sigma_x(p)$ just provides the exchange corrections to this forward scattering. Applying this theory to calculating the self-energies of quasiparticles in liquid ^{3}He gives

$$\Sigma_H + \Sigma_x = \frac{1}{\nu}\sum_{\mathbf{p}'\sigma'} M_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}'; \mathbf{p}, \mathbf{p}')n_{\mathbf{p}'\sigma'} \tag{11.189}$$

The potential $V(0)$ has been replaced by the forward scattering matrix element $M_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}'; \mathbf{p}, \mathbf{p}')$, and the distribution of quasiparticles is written as $n_{\mathbf{p}'\sigma'}$. When an excitation is created, the change in this self-energy function due to the excitation is

$$\frac{1}{\nu}\sum_{\mathbf{p}'\sigma'} M_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}'; \mathbf{p}, \mathbf{p}')\delta n_{\mathbf{p}'\sigma'} \tag{11.190}$$

This expression is very familiar, since it is exactly the form of the interaction energy in Fermi liquid theory, where $f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'} = M_{\sigma\sigma'}(\mathbf{p}, \mathbf{p}'; \mathbf{p}, \mathbf{p}')$. The Fermi liquid parameters represent the forward scattering amplitude between bare quasiparticles. It will be necessary to obtain the scattering amplitude for other directions besides forward and also to have it for dressed quasiparticles of energy $\bar{\varepsilon}_{\mathbf{p}}$ rather than bare ones of energy $\varepsilon_{\mathbf{p}}$.

The objective is to have a model for the scattering of two quasiparticles. The parameters of the model are chosen to reproduce the Fermi liquid theory for scattering in the forward direction. One method is to write an effective quasiparticle Hamiltonian with creation and destruction operators:

$$H = \sum_{\mathbf{p}\sigma}(\varepsilon_{\mathbf{p}} - \mu)C^\dagger_{\mathbf{p}\sigma}C_{\mathbf{p}\sigma} + \tfrac{1}{2}\sum_{\mathbf{pkq}}\sum_{ss'}\bar{M}(\mathbf{q})C^\dagger_{\mathbf{p}+\mathbf{q},s}C^\dagger_{\mathbf{k}-\mathbf{q},s'}C_{\mathbf{k}s'}C_{\mathbf{p}s} \tag{11.191}$$

The matrix element $\bar{M}$ is different from $M_{\sigma\sigma'}$, because $\bar{M}(\mathbf{q})$ does not include exchange events. In the theory of the homogeneous electron gas, which has a similar Hamiltonian, the interactions between particles led to a screening of the interactions. The word screening sometimes implies a charge redistribution, so the alternate word dressing is used in liquid ^{3}He. But the physics of the two concepts is the same: the quasiparticles rearrange all the other quasiparticles in their vicinity to alter the effective interaction between any pair of them. In the random phase approximation, this effective interaction is

$$M_{\text{eff}}(\mathbf{q}, \omega) = \frac{\bar{M}(\mathbf{q})}{1 - \bar{M}(\mathbf{q})P^{(1)}(\mathbf{q}, \omega)} \tag{11.192}$$

where $P^{(1)}$ is the simple bubble contribution for the polarization diagram. The denominator has the dressing function $1 - MP^{(1)}$, which is equivalent to the dielectric screening function in the electron gas. the forward direction is obtained by first taking $\omega \to 0$, and afterwards $\mathbf{q} \to 0$:

$$\lim_{\omega,\mathbf{q}\to 0}\begin{cases} P^{(1)} = -N_F \\ \bar{M}(\mathbf{q}) = F^s_0/N_F \\ N_F M_{\text{eff}} = \dfrac{N_F\bar{M}(0)}{1 + N_F\bar{M}(0)} = \dfrac{F^s_0}{1 + F^s_0} \equiv A^s_0 \end{cases} \tag{11.193}$$

The effective interaction F^s_0 is replaced by A^s_0. The quantity F^s_0 is the s-wave part of the interaction between bare quasiparticles, while A^s_0 is the same interaction between dressed quasiparticles. This difference is the same as between the bare Coulomb interaction e^2/r and the screened interaction. The dressed interaction between quasiparticles is changed, from the

bare interaction, by the other quasiparticles which have also been influenced to alter their equilibrium configuration by the effective interaction between quasiparticles.

Since the static limit of (11.193) has been obtained, the next step is to examine the dynamical predictions of this model. Consider the case whereby $\omega \gg qv_F$. For the electron gas in this same limit recall that

$$\lim_{\omega \gg qv_F} \begin{cases} P^{(1)}(\mathbf{q}, \omega) = \dfrac{nq^2}{m^* \omega^2} \\ \bar{M}(\mathbf{q}) P^{(1)}(\mathbf{q}, \omega) = \dfrac{F_0^s}{3} \dfrac{k_F^2 q^2}{m^{*2} \omega^2} \equiv \dfrac{c_1^2 q^2}{\omega^2} \\ c_1^2 = \dfrac{F_0^s k_F^2}{3m^{*2}} = \dfrac{v_{F0}^2}{3} \dfrac{F_0^s}{1 + F_1^s/3} \end{cases} \tag{11.194}$$

This result may be used to obtain the dressed susceptibility, which is the correlation function for density–density operators:

$$\chi_d(\mathbf{q}, i\omega) = -\int_0^\beta d\tau e^{i\omega_n \tau} \langle T_\tau \rho(\mathbf{q}, \tau) \rho(-\mathbf{q}, 0) \rangle \tag{11.195}$$

$$\lim_{\omega \gg qv_F} N_F \chi_d(\mathbf{q}, \omega) = \frac{N_F \bar{M}(\mathbf{q})}{1 - N_F \bar{M}(\mathbf{q})} = \frac{c_1^2 q^2}{\omega^2 - c_1^2 q^2} \tag{11.196}$$

There is a resonance at $\omega \approx c_1 q$. The large value of F_0^s ensures that $c_1 > v_F$ so $\omega > v_F q$ is valid when $\omega \approx c_1 q$.

Table 11.2 shows that the exact sound velocity in the long-wavelength limit is $c^2 = (v_{F0}^2/3)(1 + F_0^s)/(1 + F_1^s/3)$. The approximate result in (11.196) is close to the exact c, since F_0^s is large and there is not much difference between F_0^s and $1 + F_0^s$. The factor of $(1 + F_1^s/3)$ comes from the effective mass. The exact density–density correlation function has the exact same form as (11.196) but with the correct speed of sound c rather than the approximate result c_1. The difference between these two calculations is that the long-wavelength sound velocity c is found in the hydrodynamic regime and is the first sound. The above calculation is for the collisionless regime, and c_1 is the velocity of zero sound. They approach the same value for large F_0^s. The density–density correlation function has the long-wavelength limit of sound waves, which dominates the excitation spectrum.

Define $A_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$ as the Fermi liquid theory function which gives the interaction between dressed quasiparticles. It is also the forward scattering amplitude of the screened matrix element for the interaction between two quasiparticles. It may be derived from the bare interaction between quasiparticles $f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$ by the following simple argument. The self-energy of a quasiparticle may be expressed as either the product of the bare interaction $f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$, and the bare additional density $\delta n_{\mathbf{p}\sigma}$, or else the product of the dressed interaction $A_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$ and the dressed additional density $\delta \bar{n}_{\mathbf{p}\sigma}$:

$$\bar{\varepsilon}_{\mathbf{p}} - \varepsilon_{\mathbf{p}} = \sum_{\mathbf{p}'\sigma'} f_{\mathbf{p}\sigma,\mathbf{p}'\sigma'} \delta n_{\mathbf{p}'\sigma'} = \sum_{\mathbf{p}'\sigma'} A_{\mathbf{p}\sigma,\mathbf{p}'\sigma'} \delta \bar{n}_{\mathbf{p}'\sigma'} \tag{11.197}$$

The interaction between dressed quasiparticles $A_{\mathbf{p}\sigma,\mathbf{p}'\sigma'}$ is also defined only right on the Fermi surface, where both $\mathbf{p}$ and $\mathbf{p}'$ are equal to k_F. All quantities are expanded in spherical

harmonics, and doing the angular integrals gives the following obvious formula relating these coefficients:

$$A_{\mathbf{p}\sigma,\mathbf{p}'\sigma'} = \frac{1}{N_F}\sum_l [A_l^s + \boldsymbol{\sigma}\cdot\boldsymbol{\sigma}' A_l^a] P_l(\cos\theta) \tag{11.198}$$

$$A_l^{s,a}\delta\bar{n}_{lm}^{s,a} = F_l^{s,a}\delta n_{lm}^{s,a} \tag{11.199}$$

Using the previous relationship (11.180) between $\delta\bar{n}_{lm}^{s,a}$ and $\delta n_{lm}^{s,a}$, the final formula is deduced between A_l and F_l

$$A_l^s = \frac{F_l^s}{1 + F_l^s/(2l+1)} \tag{11.200}$$

$$A_l^a = \frac{F_l^a}{1 + F_l^a/(2l+1)} \tag{11.201}$$

The case $l = 0$ agrees with the RPA result obtained in (11.193). These numbers are tabulated in Table 11.3. The effective interaction between dressed quasiparticles is qualitatively much different than the interaction between bare quasiparticles. The very large value of F_0^s makes A_0^s less than unity, so the dressing effects have a drastic effect on reducing the effective interaction. Of course, the same thing happens for electron–electron interactions in metals, where screening makes a similar reduction. But the most interesting effect is the increase in the spin-dependent parts A_l^a. These become larger in magnitude than F_l^a because the latter are negative. This enhancement of the spin-dependent interaction is one step in the argument, which is given later, toward the conclusion that superfluidity in liquid ^{3}He is caused by triplet pairing.

The next question is the method of calculating the spin susceptibility. In (11.194) and (11.196), RPA and Fermi liquid parameters were used to derive an approximate formula for the density–density correlation function. Is it possible to do the same for the spin susceptibility? By reasoning deductively, Leggett (1965) suggested that a possible form for the spin susceptibility in the RPA is

$$\chi_s(\mathbf{q},\omega) = \frac{P^{(1)}(\mathbf{q},\omega)}{1 - (F_0^a/N_F)P^{(1)}(\mathbf{q},\omega)} \tag{11.202}$$

This equation is deduced, in analogy with (11.192) and (11.193), by just replacing F_0^s by F_0^a. A better treatment would start from a model Hamiltonian which would give this result for the correlation function. For the moment, accept (11.202) as a reasonable hypothesis and examine its dynamical predictions.

The spectral function for this retarded correlation function is called $S_p(\mathbf{q},\omega)$. In analogy with the electron gas, two types of dynamical modes are expected from this correlation function. The first is a collective mode equivalent to the plasmon in the electron gas and the sound waves in the density–density correlation function. Such modes can exist only when $\omega \gg qv_F$. The theory has already been done for the sound case, in (11.194) and (11.196), so change F_0^s to F_0^a in these equations. The prediction is that the collective *spin sound* has the eigenfrequency

$$\omega = qv_F\sqrt{\frac{F_0^a}{3}} \tag{11.203}$$

This frequency is imaginary, since F_0^a is negative. These modes are totally damped and do not exist as excitations.

The other excitation of the fermion gas is the particle–hole pairs. The theory of *spin pairs* can be developed in an analogous fashion. Consider the case where $\omega < qv_F$ and $q < k_F$. Then the polarization function $P^{(1)}$ can be approximated for its real and imaginary parts by

$$P^{(1)}(\mathbf{q}, \omega) = -N_F\left[1 - \frac{i\pi}{2}\left(\frac{\omega}{qv_F}\right)\Theta(qv_F - \omega)\right]$$
$$S_s(\mathbf{q}, \omega) = -2\,\mathrm{Im}[\chi_s(\mathbf{q}, \omega)] = \lambda_s \frac{\omega}{\omega^2 + \omega_a^2}\Theta(qv_F - \omega) \tag{11.204}$$

$$\omega_a = \frac{2}{\pi}\frac{qv_F}{|A_0^a|} \tag{11.205}$$

$$\lambda_a = \frac{2N_F\omega_a}{|F_0^a(1 + F_0^a)|} \tag{11.206}$$

There are no collective modes since the pole in the denominator is found at imaginary frequency. The spectral function is linear in energy for small values of $E = \hbar\omega$. The same linear behavior was found for the electron gas when the pair contribution to $\mathrm{Im}[1/\varepsilon(\mathbf{q}, \omega)]$ was evaluated, as in Fig. 5.12. The same theory can be applied to the pair spectrum from the density–density correlation function, and the only difference is in replacing F_0^a by F_0^s and A_0^a by A_0^s. It is instructive to compare these two cases when pairs are made by either the spin or density response functions. The energy widths ω_a and ω_s are similar for the two cases, since A_0^a and A_0^s differ only by a factor of 2. However, the significant difference is in the coupling strength $\lambda_{s,a}$. The factor $|F_0^{s,a}(1 + F_0^{s,a})|^{-1}$ is quite different for the two cases, since it is 4.5 for spins and 0.009 for density, which differ by a factor of 500. The RPA predicts that the spectral function for making pairs from the spin correlation function is 10^3 larger than for the density correlation function.

The correlation function (11.202) must be regarded as *ad hoc* without a derivation based on a realistic model Hamiltonian for the interaction between quasiparticles. The only attempt in this direction has been using *paramagnon theory*, which is briefly described. This theory starts with a model Hamiltonian for quasiparticles of the form (11.191), with the matrix element $\bar{M}$ only scattering particles of opposite spin. This matrix element is taken to be independent of momentum and is denoted by the symbol I:

$$V = I\int n_\uparrow(\mathbf{r})n_\downarrow(\mathbf{r})d^3r \tag{11.207}$$

$$= \frac{I}{v}\sum_{\mathbf{kk'q}} C^\dagger_{\mathbf{k+q}\uparrow}C_{\mathbf{k}\uparrow}C^\dagger_{\mathbf{p-q}\downarrow}C_{\mathbf{p}\downarrow} \tag{11.208}$$

This interaction term was first suggested by Berk and Schrieffer (1966) in their discussion of the superconductivity effects in transition metals, such as Pd, which are nearly ferromagnetic. They are actually paramagnetic, with a large spin susceptibility and a low transition temperature for superconductivity. They reasoned that the spin arrangements in Pd have short-range correlation, whereby if any spin was pointing in a direction, then nearby spins were highly correlated and were likely pointing in the same direction. The spin arrangements were ordered locally, although for short durations, and these spin fluctuations are called paramagnons. They proposed that these spin fluctuations suppressed the onset of superconductivity, since an electron with the spin pointing in the opposite direction would find it hard to be in the same vicinity because of Coulomb repulsion. They constructed this ad hoc Hamiltonian, which had a point repulsion ($I > 0$) between particles of opposite spin.

This idea was adopted immediately for liquid ^{3}He by Doniach and Engelsberg (1966) and Rice (1967). They used it to calculate a number of properties of liquid ^{3}He. Levin and Valls (1978) show that paramagnon theory gives an accurate quantitative theory of dressed quasiparticle interaction for both superfluidity and quasiparticle transport.

An alert reader will notice that the above Hamiltonian is identical to the Hubbard model which is discussed in Chapter 6. Of course, the Hubbard model was for a lattice, while here it is for a liquid. In the wave vector representation, the two theories are identical. The reader can refer back to Chapter 6 for the predictions of this model Hamiltonian.

11.2.4. Quasiparticle Transport

Fermi liquid theory can also be used to describe the transport of quasiparticles at very low temperatures. The agreement between theory and experiment provides a further verification of Fermi liquid theory and the derived parameters. The theoretical framework was provided by Abrikosov and Khalatnikov (1959), who presented an approximate theory of viscosity η and thermal conductivity K. The third transport coefficient is spin diffusion D, and an approximate theory was first given by Hone (1961). All these theories involve the solution of the Boltzmann equation for transport in Fermi systems, where the scattering mechanism is particle–particle interactions. In this case the particles are quasiparticles. Later there was an exact solution of the Boltzmann equation in the limit where $T \to 0$ by Jensen *et al.* (1968) and Brooker and Sykes (1968). It would take us too far afield to solve the Boltzmann equation for these three transport coefficients, so the results are just quoted. The following formulas are exact in the limit of zero temperature:

$$K = \frac{8\pi^2 k_F^3}{3T(m^*)^4}\left[\left\langle \frac{W(\theta,\phi)(1-\cos\theta)}{\cos(\theta/2)}\right\rangle\right]^{-1} H(\lambda_K)$$

$$\eta = \frac{64 k_F^5}{45(m^*)^4(k_B T)^2}\left[\left\langle \frac{W(\theta,\phi)(1-\cos\theta)^2\sin^2\phi}{\cos(\theta/2)}\right\rangle\right]^{-1} C(\lambda_\eta)$$

$$D = \frac{32\pi^2 k_F^2(1+F_0^a)}{3(m^*)^5(k_B T)^2}\left[\left\langle \frac{W_{\uparrow\downarrow}(\theta,\phi)(1-\cos\theta)(1-\cos\phi)}{\cos(\theta/2)}\right\rangle\right]^{-1} C(\lambda_D)$$

$$H(\lambda) = \frac{3-\lambda}{4}\sum_{n=0}^{\infty}\frac{4n+5}{(n+1)(2n+3)[(n+1)(2n+3)-\lambda]} \tag{11.209}$$

$$C(\lambda) = \frac{1-\lambda}{4}\sum_{n=0}^{\infty}\frac{4n+3}{(n+1)(2n+1)[(n+1)(2n+1)-\lambda]}$$

$$\lambda_K = 3 - 2\frac{\langle W(\theta,\phi)(1-\cos\theta)/\cos(\theta/2)\rangle}{\langle W(\theta,\phi)/\cos(\theta/2)\rangle}$$

$$\lambda_\eta = 1 - \frac{3}{4}\frac{\langle W(\theta,\phi)(1-\cos\theta)^2\sin^2\phi/\cos(\theta/2)\rangle}{\langle W(\theta,\phi)/\cos(\theta/2)\rangle}$$

$$\lambda_D = 1 - \frac{1}{2}\frac{\langle W_{\uparrow\downarrow}(\theta,\phi)(1-\cos\theta)^2\sin^2\phi/\cos(\theta/2)\rangle}{\langle W(\theta,\phi)/\cos(\theta/2)\rangle}$$

The functions $H(\lambda)$ and $C(\lambda)$ are correction factors. The original solution of Abrikosov, Khalatnikov, and Hone had an approximate result with these quantities set equal to unity. In the exact solution, one also has to evaluate these correction terms. The first step is to find the appropriate λ and then to use it in the series for $H(\lambda)$ or $C(\lambda)$. Both series converge rapidly, so that the answer is obtained easily. The factor C turns out to be about 0.8, so that the older theories are reduced by about 20%. However, the correction for the thermal conductivity is much larger, since $H(\lambda_K)$ is almost exactly 0.5, so the older expression erred by a factor of 2. These correction factors, resulting from the exact solution, play an important role in improving the agreement between theory and experiment.

These formulas contain one or several brackets which have a factor $W(\theta, \phi)$ plus some angular terms. This average is taken over the 4π solid angle:

$$\left\langle W \frac{1-\cos\theta}{\cos(\theta/2)} \right\rangle = \frac{1}{4\pi}\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin(\theta) W(\theta, \phi) \frac{1-\cos\theta}{\cos(\theta/2)} \tag{11.210}$$

The angles θ and ϕ are explained later. The symbol $W(\theta, \phi)$ is the matrix element for the scattering of quasiparticles. It is conventionally defined as

$$\begin{aligned} W(\theta, \phi) &= \frac{\pi}{\hbar}\left[|A_{\uparrow\downarrow}(\theta, \phi)|^2 + \tfrac{1}{2}|A_{\uparrow\uparrow}(\theta, \phi)|^2\right] \\ W_{\uparrow\downarrow}(\theta, \phi) &= \frac{2\pi}{\hbar}|A_{\uparrow\downarrow}(\theta, \phi)|^2 \end{aligned} \tag{11.211}$$

where $A_{\uparrow\uparrow}$ and $A_{\uparrow\downarrow}$ are the matrix elements for the scattering of two spin-up particles, and a spin-up with a spin-down particle. These matrix elements are given by the Fermi liquid parameters. The factor $(\frac{1}{2})$ in front of $|A_{\uparrow\uparrow}|^2$ occurs because the scattering of identical quasiparticles into $(\mathbf{k}, -\mathbf{k})$ is indistinguishable from $(-\mathbf{k}, \mathbf{k})$, and the angular average counts both of these and so overcounts the scattering by a factor of 2.

The factors in the numerators of (11.209) come from the effective lifetime appropriate to the transport coefficient. For the thermal conductivity, the effective lifetime contains the factor $(1 - \cos\theta)$, just as it does for the electrical conductivity in metals. The other factor in this angular average is $1/\cos(\theta/2)$, whose origin is now described. First examine the expression for the lifetime of a quasiparticle as it scatters from the other quasiparticles. Assume that the initial quasiparticle has momentum $\mathbf{p}_1$, that it scatters from a quasiparticle $\mathbf{p}_2$, and that they go to $\mathbf{p}_3$ and $\mathbf{p}_4$. The lifetime for this scattering rate is written in a symmetical form

$$\begin{aligned} \frac{1}{\tau(\mathbf{p}_1)} = \frac{2\pi}{\hbar}\int \frac{d^3p_2 d^3p_3 d^3p_4}{(2\pi)^9} \sum_{\sigma'} |A_{\uparrow\sigma'}|^2 (2\pi)^3 \delta(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_3 - \mathbf{p}_4) \\ \times \delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 - \varepsilon_4) n_2 (1 - n_3)(1 - n_4) \end{aligned} \tag{11.212}$$

The occupation factors $n_j = n_F(\varepsilon_j)$ express the feature that $\mathbf{p}_2$ is in an occupied state, while $\mathbf{p}_3$ and $\mathbf{p}_4$ must go to empty ones. The integrand contains the delta functions for conservation of energy and momentum, and the matrix element $|A_{\uparrow\sigma'}|^2$ for scattering (but $|A_{\uparrow\uparrow}|^2$ is divided by two). The evaluation of this quantity follows Pines and Nozieres (1966).

At very low temperature, all the quasiparticles will be quite close to the Fermi surface. The magnitude of all the momentums $\mathbf{p}_j (j = 1, 2, 3, 4)$ are very near to k_F and therefore very nearly the same length. This feature simplifies the calculation. The delta function for momentum conservation is used to eliminate the d^3p_4 integral and leaves the combination $d^3p_2 d^3p_3$. Let $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2$ be the center of mass momentum of the two quasiparticles, which

is unchanged by the collision. The d^3p_3 integral is nested inside d^3p_2 so that the vector $\mathbf{P}$ is used as the basis of choosing the angles in doing the d^3p_3 integral. In particular, let $\mathbf{P}\cdot\mathbf{p}_3 = Pp_3\cos(\theta_3)$, $\cos(\theta_3) \equiv v_3$. Then the vector $\mathbf{p}_4$ has the magnitude

$$p_4 = |\mathbf{P} - \mathbf{p}_3| = \sqrt{P^2 + p_3^2 - 2Pp_3v_3} \tag{11.213}$$

This identity is used to change the variables of integration from $d\theta_3\sin(\theta_3) = dv_3$ to dp_4, where $p_3dv_3 = -p_4dp_4/P$. The next step is to define P in terms of the variables used in the d^3p_2 integral. Use θ to denote the angle between $\mathbf{p}_1$ and $\mathbf{p}_2$ and use the fact that both are near k_F to find $P = |\mathbf{p}_1 + \mathbf{p}_2| = \sqrt{p_1^2 + p_2^2 + 2p_1p_2\cos(\theta)} \approx k_F\sqrt{2(1+\cos\theta)} = 2k_F\cos(\theta/2)$. Finally, we set $p_2^2dp_2 \approx k_Fd(p_2^2)/2$ which brings us to the identity

$$d^3p_2d^3p_3 = (m^*)^3\frac{\sin(\theta)d\theta}{2|\cos(\theta/2)|}d\phi_2d\phi_3d\varepsilon_2d\varepsilon_3d\varepsilon_4 \tag{11.214}$$

The six integration variables in $d^3p_2d^3p_3$ have been reexpressed in terms of coordinates which are more useful. It is usually assumed that the matrix element has no significant energy variation. Then the energy integrals can be done exactly. The first one is trivial, since it just uses the delta function for energy conservation. The remaining two integrals can also be done exactly, which is assisted by the indicated variable changes:

$$I = \int d\varepsilon_2d\varepsilon_3d\varepsilon_4\delta(\varepsilon_1 + \varepsilon_2 - \varepsilon_3 - \varepsilon_4)n_F(\varepsilon_2)n_F(-\varepsilon_3)n_F(-\varepsilon_4) \tag{11.215}$$

$$I = \int d\varepsilon_2d\varepsilon_3n_F(\varepsilon_2)n_F(-\varepsilon_3)n_F(\varepsilon_3 - \varepsilon_1 - \varepsilon_2) \tag{11.216}$$

$$x = e^{\beta(\varepsilon_2-\mu)}, \qquad y = e^{\beta(\varepsilon_3-\mu)}, \qquad z = e^{-\beta(\varepsilon_1-\mu)} \tag{11.217}$$

$$I = (k_BT)^2\int_0^\infty dx\int_0^\infty dy\frac{1}{(x+1)(y+1)(x+yz)} \tag{11.218}$$

$$= (k_BT)^2\int_0^\infty\frac{dy}{y+1}\frac{\ln(yz)}{yz-1} \tag{11.219}$$

$$= \frac{1}{2\beta^2}\left(\frac{\pi^2 + \ln^2 z}{z+1}\right) = \tfrac{1}{2}(1-n_1)[\pi^2(k_BT)^2 + (\varepsilon_1-\mu)^2] \tag{11.220}$$

The last integral is from G&R, 4.232(3). The form of the answer is very interesting. The result is proportional to two terms: one is $(\pi k_BT)^2$ and the other is $(\varepsilon_1 - \mu)^2$. The latter vanishes at the chemical potential $\varepsilon_1 = \mu$ and the former vanishes at zero temperature. This form of the answer is found in particle–particle scattering of any like fermions. The same factor enters into the expression for the inverse lifetime of an electron in a metal, from electron–electron scattering.

The last step is to do the angular integrals. First it is helpful to visualize the coordinate system. Take the case where all four vectors $\mathbf{p}_j$ have the same length k_F. Define θ as the angle between $\mathbf{p}_1$ and $\mathbf{p}_2$. It is also the angle between $\mathbf{p}_3$ and $\mathbf{p}_4$, which is true only when all four vectors are the same length. Only one more angle is needed to define the relative orientation of all four vectors. This angle could be the relative orientation of $\mathbf{p}_1$ and $\mathbf{p}_3$, but this choice is not conventional. Instead, note that the two initial vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ form a plane whose orientation is represented by a unit vector $\hat{n}_i$ in the direction $\mathbf{p}_1\times\mathbf{p}_2$. Similarly, the scattering plane in the final state has an orientation $\hat{n}_f$ in the direction $\mathbf{p}_3\times\mathbf{p}_4$. Define ϕ as the angle between these unit vectors, $\cos\phi = \hat{n}_i\cdot\hat{n}_f$. The angle ϕ is shown in the vector diagram in Fig.

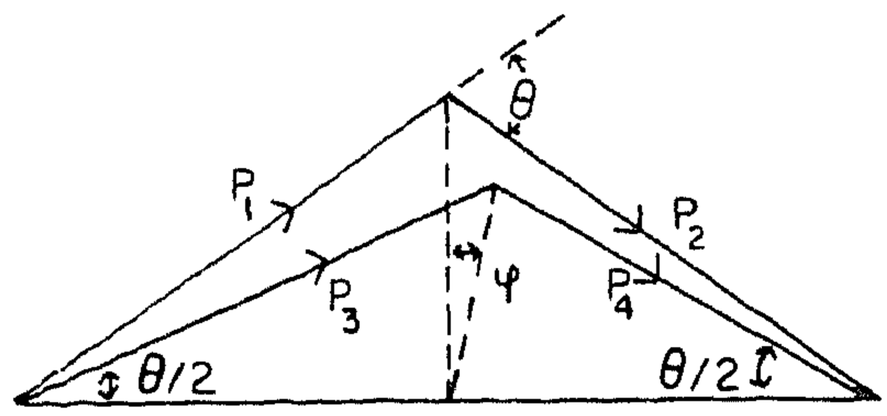


FIGURE 11.14 Angular variables in quasi particle scattering in liquid ^{3}He.

11.14. All four vectors $\mathbf{p}_j$ make an angle $\theta/2$ with respect to the center of mass momenta $\mathbf{P}$, and ϕ is the angle between the two scattering planes.

In the variable list in (11.214), the azimuthal angle ϕ_3 is actually ϕ, while ϕ_2 is unnecessary and may be integrated to give 2π. The description of the angles in the scattering process is now complete . The final lifetime is the following expression from Pines and Nozieres (1966):

$$\frac{1}{\tau(\mathbf{p}_1)} = \frac{(m^*)^3}{16\pi^4\hbar^6}(1-n_1)[\pi^2(k_BT)^2 + (\varepsilon_1-\mu)^2]\left\langle\frac{W}{\cos(\theta/2)}\right\rangle \tag{11.221}$$

The symbol $W(\theta, \phi)$ has been given its definition in (11.211), where all the angular variables have been defined explicitly. The result for $1/\tau(\mathbf{p}_1)$ has two terms which show that at the chemical potential $\varepsilon_1 = \mu$ the lifetime is determined by the temperature, but at high energies $\beta(\varepsilon_1 - \mu) \gg 1$ it is determined by phase space availability. The equivalent quantity using Green's functions is $-2\,\mathrm{Im}[\Sigma(\mathbf{p}_1, \xi_1)]$ for the appropriate diagram. The imaginary part of the self-energy produces the same lifetime if the diagrams are selected correctly, except the $(1 - n_1)$ factor is absent (see Problem 7 at the end of this chapter). The factor $(1 - n_1)$ should not be included in the quasiparticle lifetime.

The evaluation of the angular integrals for the quasiparticle lifetime require an expression for $W(\theta, \phi)$. The Landau–Fermi liquid theory gives only the forward part of the scattering amplitude, where $\mathbf{p}_3 = \mathbf{p}_1$, $\mathbf{p}_4 = \mathbf{p}_2$, or the opposite. Forward scattering corresponds to the case where $\phi = 0$ or π. A method is needed of extending this scattering theory to other values of angle besides the forward direction. The extension to other angles is useful not only for this calculation of the quasiparticle lifetime but also for the similar calculations of the transport coefficients in (11.209) which have other angular averages of $W(\theta, \phi)$. In transport theory, the quasiparticles not only scatter in the forward direction but to any other point on the Fermi surface. How can this be described by using only known parameters?

An ingenious and very successful method for this was proposed by Dy and Pethick (1969), which they called the s-p approximation. The goal is to obtain the scattering amplitudes $A_{\uparrow\uparrow}(\theta, \phi)$ and $A_{\uparrow\downarrow}(\theta, \phi)$ for the scattering of two spin-up particles and a spin-up with a spin-down particle. If two particles both have spin up, they must be in a relative triplet spin state, and this scattering amplitude is labeled $A_t(\theta, \phi)$. The singlet scattering amplitude is called $A_s(\theta, \phi)$. Two particles in opposite spin states have equal likelihood of being in a singlet or triplet state:

$$A_{\uparrow\uparrow}(\theta, \phi) = A_{\downarrow\downarrow}(\theta, \phi) = A_t(\theta, \phi) \tag{11.222}$$

$$A_{\uparrow\downarrow}(\theta, \phi) = A_{\downarrow\uparrow}(\theta, \phi) = \tfrac{1}{2}[A_t(\theta, \phi) + A_s(\theta, \phi)] \tag{11.223}$$

The next step is to use the symmetry of the two-particle wave function to deduce the possible angular variations. In center of mass coordinates, the only relevant angular variable in the

scattering process is that between the initial and final relative momenta, which is ϕ. Singlet states must be symmetric under the exchange of the orbital part of the particle coordinates, so only even values of relative angular momentum are allowed, as explained in Sec. 11.1. Similarly, triplet spin states permit only odd values of relative angular momentum, so that the symmetry of the two-particle wave functions dictates the following possible choices for these scattering amplitudes:

$$A_t(\theta, \phi) = \sum_{l\,\text{odd}} C_l(\theta) P_l(\cos\phi) \tag{11.224}$$

$$A_s(\theta, \phi) = \sum_{l\,\text{even}} C'_l(\theta) P_l(\cos\phi) \tag{11.225}$$

The next step is the s-p approximation. Dy and Pethick argue that since only the $l = 0$ and 1 values are known in the spherical harmonic expansion for θ, consistency dictates that only similar terms are retained in the expansion for Legendre functions on the ϕ variable. In this case there is only one term allowed in each expansion, which is $l = 0$ for singlet states and $l = 1$ for triplet states:

$$A_t(\theta, \phi) = C_1(\theta)\cos(\phi) \tag{11.226}$$

$$A_s(\theta, \phi) = C'_0(\theta) \tag{11.227}$$

The next step in the derivation is to find the amplitudes $C_1(\theta)$ and $C'_0(\theta)$ by evaluating them in the forward direction, where the answer is known. Set $\phi = 0$ and equate these coefficients to the interaction term between dressed quasiparticles:

$$C_1(\theta) = A_t(\theta, 0) = \frac{1}{N_F}\sum_{l=0}^{1}(A_l^s + A_l^a)P_l(\cos\theta) \tag{11.228}$$

$$C'_0(\theta) = A_s(\theta, 0) = \frac{1}{N_F}\sum_{l=0}^{1}(A_l^s - 3A_l^a)P_l(\cos\theta) \tag{11.229}$$

These various steps are collected together to finally provide the angular variation for the scattering amplitude between dressed quasiparticles:

$$A_{\uparrow\uparrow}(\theta, \phi) = \frac{\cos\phi}{N_F}[(A_0^s + A_0^a) + (A_1^s + A_1^a)\cos\theta] \tag{11.230}$$

$$\begin{aligned} A_{\uparrow\downarrow}(\theta, \phi) = \frac{1}{2N_F}\{&\cos\phi[(A_0^s + A_0^a) + (A_1^s + A_1^a)\cos\theta] \\ &+ [(A_0^s - 3A_0^a) + (A_1^s - 3A_1^a)\cos\theta]\} \end{aligned} \tag{11.231}$$

In older calculations of the transport coefficients, the angular functions $A_{\sigma\sigma'}(\theta, \phi)$ were approximated by the results in the forward direction $\phi = 0$, since they are given by the Landau Fermi liquid parameters. The improvement of Dy and Pethick is to add the factor $\cos\phi$ to the amplitude for the triplet scattering. The factor of $\cos\phi$ is a necessary addition from the point of symmetry, since the triplet amplitude must change sign when $\phi = \pi$, because of the wave function antisymmetry. Going from $\phi = 0$ to $\phi = \pi$ effectively changes the scattering ($\mathbf{p}_1 \to \mathbf{p}_3, \mathbf{p}_2 \to \mathbf{p}_4$) to the exchange event ($\mathbf{p}_1 \to \mathbf{p}_4, \mathbf{p}_2 \to \mathbf{p}_3$). This small change of adding $\cos\phi$ is a big one insofar as improving the agreement between theory and experiment for the transport coefficients. Table 11.4 shows their calculation of the quantities KT, DT^2, and ηT^2, as compared with the experimental values, using the theoretical formulas in (11.209). The agreement is obviously excellent. Previous calculations without the $\cos\phi$

TABLE 11.4 ^{3}He transport results as $T \to 0$

	KT (erg/cm-s)	DT^2 (cm^2-K^2/s)	ηT^2 (poise-K^2)
Experiment[a]	35	1.4×10^{-6}	1.8×10^{-6}
Theory[b]	33	1.6×10^{-6}	1.6×10^{-6}

a. Wheatley (1975).
b. Dy and Pethick (1969).

factor in A_t disagreed with experiment by a factor of two. Dy and Pethick also predicted the transport results for higher temperatures.

Another important success of the s-p approximation has been to explain the superfluid properties of liquid ^{3}He. In particular, it explains why the BCS type of pairing is not in the singlet state but rather is in the spin triplet state. Here the discussion follows Patton and Zaringhalam (1975). They calculated the temperature at which the Cooper instability occurs in liquid ^{3}He, which is a property of the normal state of the liquid. The construction of a wave function for the superfluid state is a different step, which is done in the next section. But the Cooper instability is a property of the normal fluid and is discussed here.

In a BCS of pairing state, two dressed quasiparticles of momentum $\mathbf{p}$ and $-\mathbf{p}$ are coupled into a collective bound state. One important question is whether this pairing occurs with the two spins in a relative singlet or triplet state. To answer this question, examine the effective interaction for each configuration. Consider the scattering of a bound pair with $(\mathbf{p}, -\mathbf{p})$ into another pair state $(\mathbf{p}', -\mathbf{p}')$. The two scattering particles have $\theta = \pi$, so that the relevant amplitudes are

$$A_t(\pi, \phi) = C_1(\pi)\hat{p}\cdot\hat{p}' \tag{11.232}$$

$$A_s(\pi, \phi) = C_0'(\pi) \tag{11.233}$$

The Fermi liquid parameters are used to derive $C_1(\pi)$ and $C_0'(\pi)$ from (11.229):

$$N_F C_1(\pi) = -2.57 \tag{11.234}$$

$$N_F C_0'(\pi) = 3.35 \tag{11.235}$$

The singlet state amplitude $C_0'(\pi)$ is positive. Two dressed quasiparticles have a repulsive interaction when paired in the spin singlet state. They do not attract each other and do not form bound states. The superfluid state is not a spin singlet in liquid ^{3}He.

For the spin triplet state, $C_1(\pi)$ is negative, which shows that the two quasiparticles have an attractive interaction. Pairing is possible in this state, and the normal fluid will show a Cooper instability at a nonzero temperature because of this tendency toward triplet pairing.

The prediction of Fermi liquid theory is that liquid ^{3}He has to form a BCS pairing state in the spin triplet arrangement. This prediction seems in good accord with the experiments on the superfluid state, which are described in the next section. The triplet pairing is a consequence of the negative value of $C_1(\pi)$ and a positive value of $C_0'(\pi)$. The physics question is: Why does this happen? The answer seems poorly understood. The parameters C_1 and C_0' are derived from the interaction between dressed quasiparticles, which in turn are derived from the interaction between bare quasiparticles. The latter numbers are taken from experiment, although the microscopic theory for them gives good results. However, the chain of argument starting at C_1 and C_0' runs back to other parameters ($F_l^{s,a}$) which are hard to calculate. It is difficult to have any intuitive insight into any of these parameters.

11.2.5. Superfluid ^{3}He

Superfluidity in liquid ^{3}He was discovered by Osheroff *et al.* (1972), who discussed it in a series of papers. Their Nobel Lectures are a nice review: Lee *et al.* (1997). They identified two phases in the liquid, which are called the A and B phases. The phase boundary on a pressure vs. temperature curve is shown in Fig. 11.15 as measured by Greywall (1986). The line marked T_c separates normal liquid ^{3}He from the superfluid phases. It is a second-order phase boundary with a jump in the specific heat, which are both characteristics of a BCS state. Both A and B phases are superfluid and are experimentally different. The line marked T_{AB} is a first-order phase boundary. The location of this phase boundary is strongly affected by small magnetic fields. Indeed, both the A and B phases have interesting magnetic properties which reinforce the notion that pairing is in a spin triplet state and that the pairs have a net magnetic moment. The theory of triplet pairing is given below.

The theory of superconductivity for electron spins in a triplet state was derived by Balian and Werthamer (1963). They developed this theory in the expectation that some metals might be superconducting with this spin configuration. The first important application of their result is in explaining superfluidity in liquid ^{3}He. Their theory is similar to the BCS theory, with weak coupling between the pairs of fermions, except that the spins are assumed to be in a relative triplet state. However, triplet pairing makes a nontrivial change in the nature of the answer because of the additional spin degrees of freedom. The spin state ($S = 1, m_s = 1, 0, -1$) can project in different ways on the relative orbital motion ($L = 1, m_l = 1, 0, -1$). The paired states have a number of possible values of total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$. The gap equation is a matrix equation. It has a number of possible solutions, and the order parameter is actually a spinor of dimensionality 7. This feature leads to a rich description of the types of motion and ordering of the superfluid. All this complexity is just to describe one of the two phases. The theory of Balian and Werthamer is believed to describe the B phase of the superfluid. The A phase is another type of triplet pairing, which is treated later.

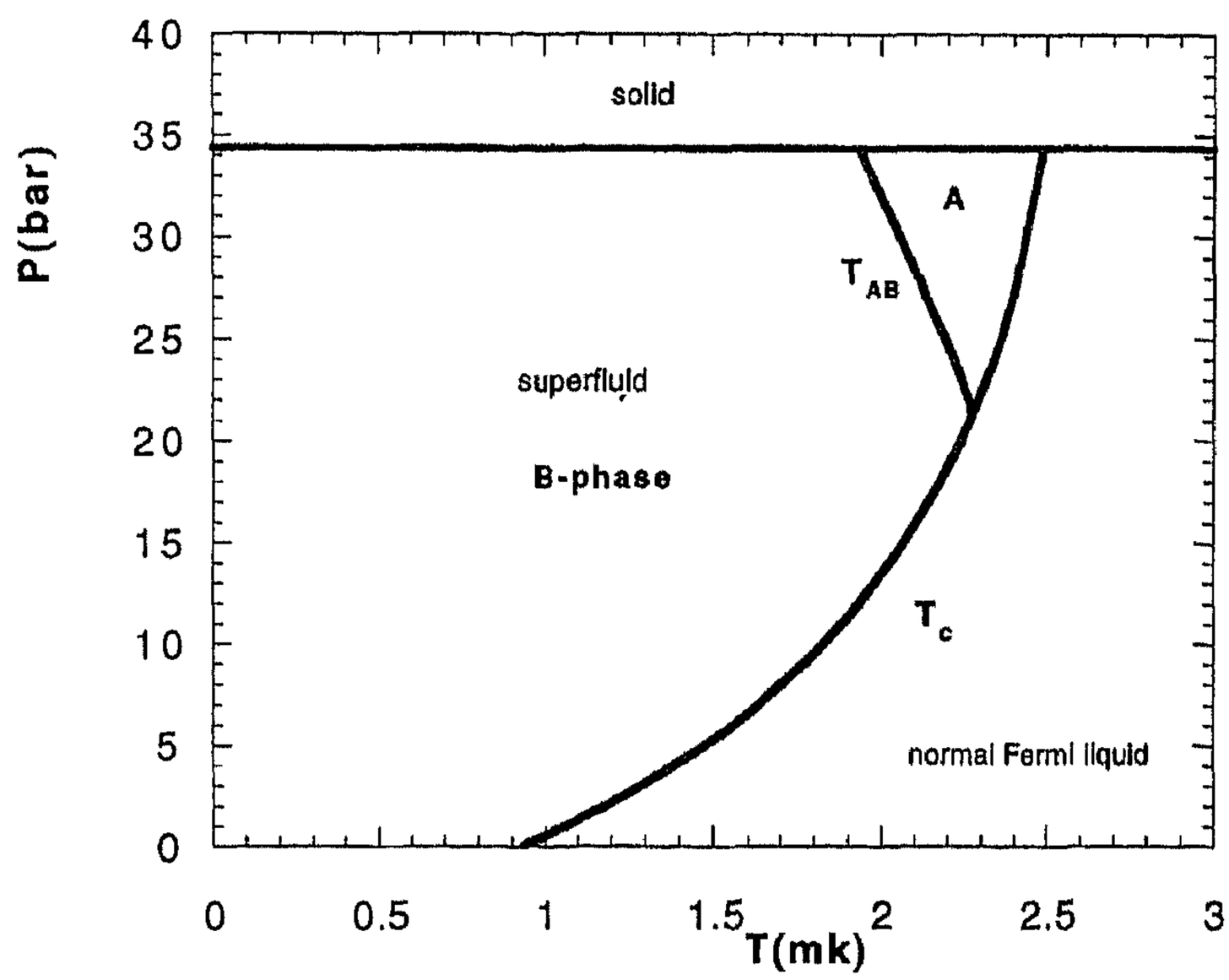


FIGURE 11.15 Phase diagram of ^{3}He as a function of pressure and temperature. The line T_c divides normal from superfluid phases. The line T_{AB} divides the two superfluid phases, and the horizontal line at 34 bars divides solid from liquid. *Source:* Greywall (1986) (used with permission)

The BW (Balian and Werthamer) theory is solved by following the same steps used to derive the BCS theory in Chapter 10. The first step is to write an effective Hamiltonian between dressed quasiparticles:

$$H = \sum_{\mathbf{p}\sigma} \xi_{\mathbf{p}} C^\dagger_{\mathbf{p}\sigma} C_{\mathbf{p}\sigma} + \frac{1}{2v} \sum_{\mathbf{p}\mathbf{p}'\mathbf{q}} \sum_{\sigma\sigma'} V(\mathbf{q}) C^\dagger_{\mathbf{p}+\mathbf{q},\sigma} C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma'} C_{\mathbf{p}'\sigma'} C_{\mathbf{p}\sigma} \tag{11.236}$$

The interaction potential $V(\mathbf{q})$ is actually a function of the momentum change between relative wave vectors $\mathbf{p}_i$ and $\mathbf{p}_f$ in the scattering process, where $\mathbf{p}_f = \mathbf{p}_i + \mathbf{q}$. The interaction potential is expanded in spherical harmonics, and the first two terms are

$$V(\mathbf{p}_f - \mathbf{p}_i) = \sum_l V_l(p_i, p_f) P_l(\hat{p}_i \cdot \hat{p}_f) = V_0 + V_1 \hat{p}_i \cdot \hat{p}_f \tag{11.237}$$

The first term V_0 acts when the orbital motion of the two particles is in a relative s state, which happens for a spin singlet. The discussion of the previous section showed the s-wave interaction is repulsive ($V_0 > 0$) and cannot cause bound states. The second term $V_1 \hat{p}_i \cdot \hat{p}_f$ is for relative p states and applies for spin triplets. This coefficient is negative and can cause pairing, where $V_1 \sim C_1(\pi)$ in the notation of the previous section. This term is the only one that is kept in the interaction potential. It has the feature that it is antisymmetric in either of the momentum variables: $V(\mathbf{p}_i, \mathbf{p}_f) = V_1 \hat{p}_i \cdot \hat{p}_f = -V(-\mathbf{p}_i, \mathbf{p}_f) = -V(\mathbf{p}_i, -\mathbf{p}_f)$. In the case of weak coupling theory, such as BCS, this potential is assumed to be a constant V_1 in energy up to a cutoff ω_c from the chemical potential. The physical nature of this cutoff is not clear. Obviously, a better theory is impossible without a better understanding of the retarded nature of the potentials between dressed or bare quasiparticles.

The next step is to define the correlation functions appropriate to the superfluid. They will be of the same type we used in the BCS theory of superconductivity. One expects superfluidity in liquid ^{3}He to be similar to a BCS state because of the obvious experimental similarities. The theory requires the introduction of seven correlation functions:

$$\begin{aligned}
\mathscr{G}(\mathbf{p}, \tau - \tau') &= -\langle T_\tau C_{\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}\sigma}(\tau') \rangle \\
\mathscr{F}_1(\mathbf{p}, \tau - \tau') &= \langle T_\tau C_{-\mathbf{p}\uparrow}(\tau) C_{\mathbf{p}\uparrow}(\tau') \rangle \\
\mathscr{F}_0(\mathbf{p}, \tau - \tau') &= \langle T_\tau C_{-\mathbf{p}\downarrow}(\tau) C_{\mathbf{p}\uparrow}(\tau') \rangle \\
\mathscr{F}_{-1}(\mathbf{p}, \tau - \tau') &= \langle T_\tau C_{-\mathbf{p}\downarrow}(\tau) C_{\mathbf{p}\downarrow}(\tau') \rangle \\
\mathscr{F}^\dagger_1(\mathbf{p}, \tau - \tau') &= \langle T_\tau C^\dagger_{\mathbf{p}\uparrow}(\tau) C^\dagger_{-\mathbf{p}\uparrow}(\tau') \rangle \\
\mathscr{F}^\dagger_0(\mathbf{p}, \tau - \tau') &= \langle T_\tau C^\dagger_{\mathbf{p}\uparrow}(\tau) C^\dagger_{-\mathbf{p}\downarrow}(\tau') \rangle \\
\mathscr{F}^\dagger_{-1}(\mathbf{p}, \tau - \tau') &= \langle T_\tau C^\dagger_{\mathbf{p}\downarrow}(\tau) C^\dagger_{-\mathbf{p}\downarrow}(\tau') \rangle
\end{aligned} \tag{11.238}$$

The quasiparticle Green's function $\mathscr{G}$ has the usual definition, although it will have a different form in the superfluid state. The others are pairing functions for different spin arrangements. The first one has both spins up and belongs to the triplet state. The next has one spin up and another down, and appears identical to the correlation function used for the BCS singlet state. However, now it is for the spin triplet state and represents the configuration where $S = 1, m_s = 0$, whereas in Chapter 10 it was the correlation function for $S = 0, m_s = 0$. The difference between these two correlation functions becomes clearer when they are written as

$$\begin{aligned}
\mathscr{F}_{S=1,m=0}(\mathbf{p}, \tau - \tau') &= -\tfrac{1}{2}[\langle T_\tau C_{-\mathbf{p}\uparrow}(\tau) C_{\mathbf{p}\downarrow}(\tau') \rangle + \langle T_\tau C_{-\mathbf{p}\downarrow}(\tau) C_{\mathbf{p}\uparrow}(\tau') \rangle] \\
\mathscr{F}_{S=0,m=0}(\mathbf{p}, \tau - \tau') &= -\tfrac{1}{2}[\langle T_\tau C_{-\mathbf{p}\uparrow}(\tau) C_{\mathbf{p}\downarrow}(\tau') \rangle - \langle T_\tau C_{-\mathbf{p}\downarrow}(\tau) C_{\mathbf{p}\uparrow}(\tau') \rangle]
\end{aligned} \tag{11.239}$$

The first one is for the spin triplet state and the second one for the spin singlet. The triplet state is antisymmetric in momentum, while the singlet state is symmetric:

$$\begin{aligned}\mathscr{F}_{S=1,m=0}(-\mathbf{p},\tau-\tau') &= -\mathscr{F}_{S=1,m=0}(\mathbf{p},\tau-\tau')\\ \mathscr{F}_{S=0,m=0}(-\mathbf{p},\tau-\tau') &= \mathscr{F}_{S=0,m=0}(\mathbf{p},\tau-\tau')\end{aligned} \tag{11.240}$$

These parity relations are easily proved from the definitions (11.239), by changing the sign on $\mathbf{p}$ and commuting the operators. The simpler definition in (11.238) can be used for either triplet or singlet with the appropriate choice of symmetry (11.240). The simple definition is used, and only the antisymmetric version is used when solving the equations self-consistently for the correlation functions of the triplet state.

There are four possible spin configurations for the two spin one-half quasiparticles: three triplets and one singlet. The most general description would have four correlation functions $\mathscr{F}_{S,m_s}$ (one for each spin arrangement) plus $\mathscr{G}$, and the gap equation becomes a 5×5 matrix equation. The problem can be simplified, at the outset, to a matrix equation which is only 3×3. The simplification is achieved by using the relationships in the previous paragraph, where $m_s=-1$ is the Hermitian conjugate of $m_s=1$ and only one $m_s=0$ function is used with momentum antisymmetry. The simplification to a matrix of dimension 3 is obviously desirable and is conventional.

The equations of motion method is used to obtain the gap equation in the weak coupling theory. The first step is to recall (10.36) for $\partial C_{\mathbf{p}\sigma}/\partial\tau$, which is then put into the τ derivative of the Green's functions:

$$\frac{\partial}{\partial\tau}C_{\mathbf{p}\sigma} = -\xi_p C_{\mathbf{p}\sigma} - \frac{1}{\nu}\sum_{\mathbf{p}'\mathbf{q}\sigma'} V(\mathbf{q})C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma'}C_{\mathbf{p}'\sigma'}C_{\mathbf{p}-\mathbf{q},\sigma} \tag{11.241}$$

$$\begin{aligned}&-\left(\frac{\partial}{\partial\tau}+\xi_p\right)\mathscr{G}(\mathbf{p},\tau-\tau') + \frac{1}{\nu}\sum_{\mathbf{p}'\mathbf{q}\sigma'} V(\mathbf{q})\\ &\times\langle T_\tau C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma'}(\tau)C_{\mathbf{p}'\sigma'}(\tau)C_{\mathbf{p}-\mathbf{q},\sigma}(\tau)C^\dagger_{\mathbf{p}\sigma}(\tau')\rangle = \delta(\tau-\tau')\end{aligned} \tag{11.242}$$

The interaction term has an expectation of four operators. These are paired in different arrangements, which produces a large number of terms, since pairings are allowed between like and unlike operators (i.e., $\langle CC\rangle, \langle C^\dagger C^\dagger\rangle, \langle CC^\dagger\rangle$) with parallel and antiparallel spin arrangements. In weak coupling theory all terms are ignored except those which contain a gap function $\mathscr{F}_m$. In the preceding equation, the terms which are retained are those with the pairing $\langle C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma'}C^\dagger_{\mathbf{p}\sigma}\rangle\langle C_{\mathbf{p}',\sigma'}C_{\mathbf{p}-\mathbf{q}\sigma}\rangle$. There are two spin arrangements, $\sigma'=\sigma$ and $\sigma'=-\sigma$, which gives the first equation of motion

$$\begin{aligned}&-\left(\frac{\partial}{\partial\tau}+\xi_p\right)\mathscr{G}(\mathbf{p},\tau-\tau') - \frac{1}{\nu}\mathscr{F}^\dagger_1(\mathbf{p},\tau-\tau')\sum_{\mathbf{q}} V(\mathbf{q})\mathscr{F}_1(\mathbf{p}-\mathbf{q},0)\\ &\qquad -\frac{1}{\nu}\mathscr{F}^\dagger_0(\mathbf{p},\tau-\tau')\sum_{\mathbf{q}} V(\mathbf{q})\mathscr{F}_0(\mathbf{p}-\mathbf{q},0) = \delta(\tau-\tau')\end{aligned} \tag{11.243}$$

Equations of motion are also needed for $\mathscr{F}_1(\mathbf{p}, \tau)$ and $\mathscr{F}_0(\mathbf{p}, \tau)$. They can both be done at the same time by considering

$$\mathscr{F}_\sigma(\mathbf{p}, \tau - \tau') = \langle T_\tau C_{-\mathbf{p}\sigma}(\tau) C_{\mathbf{p}\uparrow}(\tau')\rangle \tag{11.244}$$

$$\left(\frac{\partial}{\partial\tau'} + \xi_p\right)\mathscr{F}_\sigma(\mathbf{p}, \tau - \tau') + \frac{1}{\nu}\sum_{\mathbf{p}'\mathbf{q}\sigma'} V(\mathbf{q})$$
$$\times \langle T_\tau C_{-\mathbf{p}\sigma}(\tau) C^\dagger_{\mathbf{p}'-\mathbf{q},\sigma'}(\tau') C_{\mathbf{p}'\sigma'}(\tau') C_{\mathbf{p}-\mathbf{q},\uparrow}(\tau')\rangle = 0 \tag{11.245}$$

where $\sigma = \pm 1$ gives the two gap functions. In the bracket of four operators, pair the first two operators to get a Green's function $\mathscr{G}(-\mathbf{p}, \tau - \tau') = \mathscr{G}(\mathbf{p}, \tau - \tau')$, which sets $\sigma = \sigma'$. The other two operators are paired into a gap function. The pairings which lead to Hartree and exchange energies such as $\langle C_{-\mathbf{p}\sigma} C_{\mathbf{p}'\sigma'}\rangle\langle C^\dagger_{\mathbf{p}'-\mathbf{q}\sigma'} C_{\mathbf{p}-\mathbf{q}\uparrow}\rangle$ are omitted in weak coupling theory. The pairing terms which are kept give the following two equations of motion:

$$\left(\frac{\partial}{\partial\tau'} + \xi_p\right)\mathscr{F}_1(\mathbf{p}, \tau - \tau') - \frac{1}{\nu}\mathscr{G}(\mathbf{p}, \tau - \tau')\sum_{\mathbf{q}} V(\mathbf{q})\mathscr{F}_1(\mathbf{p} - \mathbf{q}, 0) = 0$$
$$\left(\frac{\partial}{\partial\tau'} + \xi_p\right)\mathscr{F}_0(\mathbf{p}, \tau - \tau') - \frac{1}{\nu}\mathscr{G}(\mathbf{p}, \tau - \tau')\sum_{\mathbf{q}} V(\mathbf{q})\mathscr{F}_0(\mathbf{p} - \mathbf{q}, 0) = 0$$

These are now Fourier-transformed in the usual way to obtain functions of Matsubara frequency. Then they become algebraic equations which can be solved. Important terms in these equations are the gap functions, which are defined as

$$\begin{aligned}\Delta_1(\mathbf{p}) &= -\frac{1}{\nu}\sum_{\mathbf{p}'} V(\mathbf{p} - \mathbf{p}')\mathscr{F}_1(\mathbf{p}', \tau = 0)\\ \Delta_0(\mathbf{p}) &= -\frac{1}{\nu}\sum_{\mathbf{p}'} V(\mathbf{p} - \mathbf{p}')\mathscr{F}_0(\mathbf{p}', \tau = 0)\end{aligned} \tag{11.246}$$

The equations of motion are

$$(ip - \xi_p)\mathscr{G}(p) + \mathscr{F}^\dagger_1(p)\Delta_1 + \mathscr{F}^\dagger_0(p)\Delta_0 = 1 \tag{11.247}$$

$$(ip + \xi_p)\mathscr{F}_m(p) + \mathscr{G}(p)\Delta_m = 0 \tag{11.248}$$

$$(ip + \xi_p)\mathscr{F}^\dagger_m(p) + \mathscr{G}(p)\Delta^\dagger_m = 0 \tag{11.249}$$

The equation is also given for the Hermitian conjugate $\mathscr{F}^\dagger_m$. Then the solution is straightforward:

$$E(\mathbf{p}) = \xi_p^2 + |\Delta_1(\mathbf{p})|^2 + |\Delta_0(\mathbf{p})|^2 \tag{11.250}$$

$$\mathscr{G}(p) = -\frac{ip + \xi_p}{p_n^2 + E(\mathbf{p})^2} \tag{11.251}$$

$$\mathscr{F}^\dagger_m(p) = \mathscr{F}_m(p) = \frac{\Delta_m}{p_n^2 + E(\mathbf{p})^2} \tag{11.252}$$

The effective gap function is $\Delta(\mathbf{p}) = \sqrt{|\Delta_1|^2 + |\Delta_0|^2}$ and the energy gap $E_g = 2\Delta$. In some respects, the BW state appears to be similar to a BCS singlet state but with an effective energy gap of $\Delta(\mathbf{p})$ and excitation energy $E(\mathbf{p})$. This similarity does not extend to magnetic phenomena, such as spin susceptibility, which are much more interesting in triplet pairing.

The last step in deriving the self-consistent theory is to obtain the equation for the gap function. The correlation functions $\mathscr{F}^\dagger_m(\mathbf{p}, \tau)$ are found for $\tau = 0$, with the familiar result

$$\mathscr{F}^\dagger_m(\mathbf{p}, \tau = 0) = \frac{1}{\beta}\sum_{ip} \mathscr{F}^\dagger_m(\mathbf{p}, ip) = \frac{\Delta_m(\mathbf{p})\tanh[\beta E(\mathbf{p})/2]}{2E(\mathbf{p})} \tag{11.253}$$

$$\Delta_m(\mathbf{p}) = -\int \frac{d^3p'}{(2\pi)^3} V(\mathbf{p}, \mathbf{p}') \frac{\Delta_m(\mathbf{p}')\tanh[\beta E(\mathbf{p}')/2]}{2E(\mathbf{p}')} \tag{11.254}$$

In the second equation, the definition (11.246) is used to derive the self-consistent equation for the gap function. This gap function must be solved for the energy gaps $\Delta_m(\mathbf{p})$

The gap equation is obviously a nonlinear matrix equation. Each term in the matrix depends on the excitation energy $E(\mathbf{p})$, which depends on the two energy gaps Δ_0 and Δ_1. The above is a set of coupled equations for Δ_0 and Δ_1. There is also the restriction that the gap function is an odd function of momentum, $\Delta_m(-\mathbf{p}) = -\Delta_m(\mathbf{p})$, which follows from the condition that $\mathscr{F}_m(\mathbf{p})$ have this property. This condition is automatically satisfied when the interaction potential is taken to be the p-wave type $V = V_1\hat{p}\cdot\hat{p}'$. For example, by changing dummy variables of integration in the gap equation from $\mathbf{p}'$ to $-\mathbf{p}'$, the equation is unchanged when both $V(\mathbf{p}, \mathbf{p}')$ and $\Delta_m(\mathbf{p}')$ are odd functions of $\mathbf{p}'$.

The interesting aspect of the gap equation in triplet pairing is that the solutions are not unique but have orbital degeneracy. There are two sources of angular momentum from each pair of particles: the spin angular momentum and the orbital angular momentum in the p state. The first question is the relative orientation of these angular momenta. If two particles were binding in a gas to form an He_2 molecule, then one could just couple these momenta with Clebsch–Gordon coefficients into all possible values of $J = 0, 1$, and 2. In the Fermi liquid, the superfluid state is a collective property, and each pair cannot be treated as individuals. Nevertheless, there is still some freedom in selecting the basis for the spin and the orbital angular momentum. This idea is best illustrated by an example.

The gap function $\Delta_m(\mathbf{p})$ must have p-wave symmetry. The obvious configuration has the spin and orbital states aligned along the same axis. The first solution described below was found by BW. For the gap function $\Delta_m(\mathbf{p})$ they chose the orbital function according to $Y_{1m}(\theta_p, \phi_p)$, where (θ_p, ϕ_p) are the angular orientation of the $\mathbf{p}$ vectors. Specifically,

$$\Delta_{\pm 1}(\mathbf{p}) = (\hat{p}_x \pm i\hat{p}_y)\Delta(p) = \sin(\theta_p)e^{\pm i\phi_p}\Delta(p) \tag{11.255}$$

$$\Delta_0(\mathbf{p}) = \hat{p}_z\Delta(p) = \cos(\theta_p)\Delta(p) \tag{11.256}$$

$$|\Delta(\mathbf{p})|^2 = |\Delta_0|^2 + |\Delta_1|^2 = \Delta(p)^2 \tag{11.257}$$

The functions $E(\mathbf{p})$ and $\Delta(\mathbf{p})$ are even functions of $\mathbf{p}$ and are scalar.

First verify that this choice satisfies the gap equation. Use the angles (θ, ϕ) for the direction of $\mathbf{p}$ and the variables (θ', ϕ') for $\mathbf{p}'$. From the law of cosines, the angle between $\mathbf{p}$ and $\mathbf{p}'$ is

$$\begin{aligned}
\hat{p}\cdot\hat{p}' &= \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi') \\
\Delta_1 &= -\frac{N_F V_1}{16\pi}\int d\Omega'\hat{p}\cdot\hat{p}'\sin(\theta')e^{i\phi'}\int d\xi'\frac{\Delta(p')}{E(p')}\tanh\left[\frac{\beta E(p')}{2}\right] \\
\Delta_0 &= -\frac{N_F V_1}{16\pi}\int d\Omega'\hat{p}\cdot\hat{p}'\cos(\theta')\int d\xi'\frac{\Delta(p')}{E(p')}\tanh\left[\frac{\beta E(p')}{2}\right]
\end{aligned} \tag{11.258}$$

The energy gap $E(p')$ does not depend on the angles. The angular integrals in the gap equation are easy to do and give ($g_t = -N_F V_1/12 = -C_1/24 > 0$)

$$\Delta_{\pm 1} = \Lambda \sin(\theta) e^{\pm i\phi} \tag{11.259}$$

$$\Delta_0 = \Lambda \cos(\theta) \tag{11.260}$$

$$\Lambda = g_t \int_{-\omega_c}^{\omega_c} d\xi' \frac{\Delta(p')}{E(p')} \tanh\left[\frac{\beta E(p')}{2}\right] \tag{11.261}$$

For each gap function $\Delta_m(\mathbf{p})$, the angular functions are the same on both sides of the equation. These can be canceled from both sides, and each equation produces the same formula for the effective gap function:

$$\Delta(p) = \Lambda = g_t \int_{-\omega_c}^{\omega_c} d\xi' \frac{\Delta(p')}{E(p')} \tanh\left[\frac{\beta E(p')}{2}\right] \tag{11.262}$$

The equation is identical to the BCS equation for the singlet gap. The coupling constant is effectively g_t. This equation is easiest to solve at zero temperature where the hyperbolic tangent is unity, $\Delta =$ constant:

$$1 = g_t \int_{-\omega_c}^{\omega_c} \frac{d\xi'}{\sqrt{\xi'^2 + \Delta^2}} \approx 2g_t \ln\left(\frac{2\omega_c}{|\Delta|}\right) \tag{11.263}$$

$$|\Delta|_{BW} = 2\omega_c \exp\left(-\frac{1}{2g_t}\right) \tag{11.264}$$

The solution is an equation of the BCS type for the gap function $|\Delta|$ at zero temperature. The variation with temperature below T_c is also similar to the BCS prediction.

This equation describes the Balian–Werthamer state. Their theory is thought to apply to the B state of superfluid liquid ^{3}He, although the theory must be modified to account for the strong coupling between quasiparticles. However, even the weak coupling theory, as just derived, makes a number of predictions which agree with the experimental findings in the superfluid B state. Some of these will now be described.

One important property is that the B state is *isotropic*, which is defined as having the energy gap $\Delta(p)$ not depend on angle in the superfluid. The best way to understand this definition is to give an example of a state which is nonisotropic. As a random choice, consider the possibility

$$\Delta_1 = \Delta_{-1} = 0 \tag{11.265}$$

$$\Delta_0(\mathbf{p}) = \sqrt{3}\cos(\theta)\Delta(p) \tag{11.266}$$

$$E(\mathbf{p}) = \sqrt{\xi_p^2 + 3\cos^2\theta\Delta(p)^2} \tag{11.267}$$

The p-wave pairing requires that $\Delta_m(\mathbf{p})$ always depend on a p-symmetry angular function. The choice in (11.266) provides an entirely self-consistent solution to the gap equation (11.253). The normalization factor $\sqrt{3}$ is chosen so that $\Delta(\mathbf{p})^2$ averages to $\Delta(p)^2$ around the Fermi surface. The energy gap is no longer isotropic but now depends on the angle θ between the spin and orbital motion. The energy gap and excitation energy will vary around the Fermi surface. A more detailed discussion of nonisotropic superfluid states is presented below. The motivation is provided by the superfluid A phase, which is thought to be a nonisotropic state.

The BW state (11.256) chose axes where the coordinate bases for the spin and orbital motions were aligned. There is no need to make this choice, and the same gap equation for

$\Delta(\mathbf{p})$ (and the same solution) are obtained if these axes are rotated with respect to each other. For example, give the orbital coordinates a 90° rotation, and the gap functions are

$$\Delta_{\pm 1} = (\hat{p}_y \pm i\hat{p}_z)\Delta(p) \tag{11.268}$$

$$\Delta_0 = \hat{p}_x \Delta(p) \tag{11.269}$$

$$|\Delta(\mathbf{p})|^2 = \Delta(p)^2 \tag{11.270}$$

The gap function is isotropic and given by the same value (11.264) as before. In weak coupling the choice of orbital coordinates does not influence the energy gap. The BW superfluid state has orbital degeneracy in that the orbital motion is not coupled to the spin motion and can proceed independently. This feature depends on the isotropic nature of the superfluid phase.

Many of the thermodynamic properties of the superfluid phase depend only on the energy gap. One example is the specific heat. These properties are exactly the same as predicted by the BCS theory for singlet pairing, which also has an isotropic energy gap. (In real metals, with anisotropic Fermi surfaces, the BCS gap can also be anisotropic, which is a complication we shall not discuss.) The BW theory predicts that the transition from the normal to the superfluid phase is a second-order phase transition and that the specific heat is discontinuous at this phase transition. These properties are observed in superfluid ^{3}He, as mentioned earlier. The interesting difference between singlet and triplet pairing is in magnetic phenomena. There are interesting nuclear magnetic resonance phenomena in the triplet states which were very important for identification of the superfluid phases and which verify the theory. These are described by Leggett (1975).

The A phase has a different type of triplet pairing states than the B phase. The latter is well described by the theory of Balian and Werthamer (BW). The states of the A phase are also solutions to the gap equation (11.253) for triplet pairing. They are called *equal spin pairing* (ESP). They are composed of only the $m_s = \pm 1$ components of the spin angular momentum and not with the $m_s = 0$ components. Now consider the theory when $\Delta_0 = 0$ while retaining the two gaps $\Delta_{\pm 1}$.

The first realization is that there is no relationship between the orbital momentum of the two gap functions Δ_1 and Δ_{-1}. The equation which is solved for each $\Delta_{\pm 1}$ involves the gap function Δ_0. If the latter is zero, there is no connection between $m = \pm 1$. The excitation energy $E_m(\mathbf{p}) = \sqrt{\xi_p^2 + |\Delta_m|^2}$ involves only one gap function. Another way to understand the decoupling of Δ_1 and Δ_{-1} is to start from the beginning and write the equations of motion. In describing the response of a spin-up particle, the only nozero terms are with the correlation of the motion of other spin-up particles. There is no correlation with spin-down particles once $\Delta_0 = 0$. The spin-up particles live in one world, and the spin-down particles live in another. The orbital motion of spin-up and spin-down particles is arbitrary. There are two independent superfluids, one of spin-up and one of spin-down, which coexist in the liquid.

The orbital motions of each superfluid need not be correlated. A consequence of the orbital independence is that the excitation energies $E_m(\mathbf{p})$ and gap functions $|\Delta_m(\mathbf{p})|^2$ no longer need to have the same values for $m = 1$ and $m = -1$. For example, a possible set of solutions are

$$\begin{aligned} \Delta_1(\mathbf{p}) &= \sqrt{\tfrac{3}{2}}(\hat{p}_x + i\hat{p}_y)\Delta(p) = \sqrt{\tfrac{3}{2}}\sin(\theta)e^{i\phi}\Delta(p) \\ \Delta_{-1}(\mathbf{p}) &= \sqrt{\tfrac{3}{2}}(\hat{p}_y + i\hat{p}_z)\Delta(p) = \sqrt{\tfrac{3}{2}}[\sin(\theta)\cos(\phi) + i\cos(\theta)]\Delta(p) \end{aligned} \tag{11.271}$$

It is obvious that $E_1(\theta, \phi) \neq E_{-1}(\theta, \phi)$ except at special angular points. The excitation and energy gaps for spin-up and spin-down particles would be quite different. Each energy gap function has the same shape relative to its axis (north and south poles), but these poles are rotated relative to each other for spin-up and spin-down particles.

The superfluid is called *unitary* if the spin-up and spin-down particles have the identical angular dependence of their energy gaps $E_1(\theta, \phi) = E_{-1}(\theta, \phi)$. ESP states are unitary when the north–south poles are aligned for the two systems. The superfluid is called *nonunitary* when the excitation energies are not identical as a function of angle. The BW state is unitary.

The A phase appears to be a unitary ESP state. The north and south poles are aligned for the two gap functions. This is an experimental conclusion, and there is no particular reason the state has to be unitary. The important question is the nature of the forces which tend to align the poles. There are no forces in the Hamiltonian (11.236) which cause alignment. The Hamiltonian neglects the very, very weak forces due to the dipole–dipole interactions between the nuclear moments. Leggett has shown that these dipolar forces will tend to align the poles. Although it is a weak effect, there is nothing to oppose it. Of course, any external perturbation such as a magnetic field will also tend to align the poles, since alignment lowers the magnetic energy. Nonunitary states are discussed in reviews by Leggett (1975) and Anderson and Brinkman (1975).

11.3. QUANTUM HALL EFFECTS

The quantum Hall effect (QHE), and the fractional quantum Hall effect (FQHE), are phenomena which occur in highly correlated electron fluids. The QHE was discovered by von Klitzing *et al.* (1980), and the FQHE was discovered by Tsui *et al.* (1982).

Electrons are allowed to move in only two dimensions using semiconductor technology to confine the motion in the third direction (z). The application of a strong magnetic field B along the z direction causes the electrons to move in periodic orbits in the xy plane, which are quantized. This quantization leads to energy levels $E_n = \hbar\omega_c(n + \frac{1}{2})$, where the cyclotron frequency is $\omega_c = eB/m^*c$, and m^* is the effective mass of the electron in the semiconductor. The center of the cyclotron orbit can be anywhere in the plane. The electrons select the centers of their orbits to be arrangements which are highly correlated. At certain densities the electron gas becomes quite rigid, or incompressible, which is the simple explanation of the QHE and FQHE.

The topic shares many attributes with superfluid helium. Perturbation theory does not work. The best wave functions are guessed, and are written down by intuition. The excitation spectrum is calculated using Bijl–Feynman theory.

11.3.1. Landau Levels

In a large magnetic field, electrons tend to have orbits which oscillate in time with an angular frequency given by the cyclotron frequency $\omega_c = eB/mc$ where B is the magnetic field. If the relaxation time for scattering is τ, then the behavior of electrons depends upon whether $\omega_c\tau$ is greater or smaller than one. If $\omega_c\tau \gg 1$ then the electrons form very periodic orbits. Their energy levels are quantized in units of $\hbar\omega_c$, which are called *Landau levels*. These orbits are simple to describe in two dimensions.

In the theory of electricity and magnetism, the potentials may be either scalar potentials $\phi(\mathbf{r})$ or vector potentials $\mathbf{A}(\mathbf{r})$. The magnetic field is introduced through a vector potential. The basic equation is $\mathbf{B} = \nabla \times \mathbf{A}$. When the magnetic field is a constant there are many different ways to choose a vector potential to satisfy this relation. If the magnetic field is in the z direction, then the following choices all give the correct magnetic field

$$\mathbf{A} = B(-y, 0, 0) \tag{11.272}$$

$$\mathbf{A} = B(0, x, 0) \tag{11.273}$$

$$\mathbf{A} = \frac{B}{2}(-y, x, 0) \tag{11.274}$$

Each different vector potential is a choice of *gauge condition*. Since the eigenvalues do not depend upon this choice of $\mathbf{A}$, they are called *gauge invariant*. The first gauge is chosen, for no particular reason, so that the Hamiltonian is

$$H = \sum_j \frac{1}{2m_j}\left[\mathbf{p}_j - \frac{e}{c}\mathbf{A}(\mathbf{r}_j)\right]^2 - \sum_j \boldsymbol{\mu}_j \cdot \mathbf{B} \tag{11.275}$$

$$= \frac{1}{2m}\sum_j \left[\left(p_{jx} + \frac{ey_jB}{c}\right)^2 + p_{jy}^2\right] - \sum_j \boldsymbol{\mu}_j \cdot \mathbf{B} \tag{11.276}$$

The electron has a spin, and a magnetic moment $\boldsymbol{\mu}_j$. The last term in the interaction is the Pauli interaction between the electron moment and the magnetic field. For electrons in free space, the Pauli term and the orbital eigenvalues have similar energies. The QHE and FQHE measurements are usually done on semiconductors with very small effective mass m^*, such as GaAs with $m^*/m = 0.07$. The effective mass should be used in calculating the cyclotron frequency $\omega_c = eB/m^*c$. The symbol m in our equations is actually m^*. The level splitting $\hbar\omega_c$ is greatly enhanced in the semiconductor because of the small effective mass. The Pauli energy is altered much less by the band structure, so that it is much smaller than $\hbar\omega_c$. Although small, it is important for discussing the experiments. It is ignored for now, in order to concentrate on the Landau term.

The Hamiltonian does not yet contain interactions between particles. In this case, the solution is found for one electron, and the many-particle ground state is composed of electrons which occupy these one-electron states. Consider the Hamiltonian for one electron,

$$H = \frac{1}{2m}\left[\left(p_x + \frac{eyB}{c}\right)^2 + p_y^2\right] \tag{11.277}$$

The potential energy term in (11.277) contains y as the only space variable. The Hamiltonian commutes with p_x so the eigenstates can be chosen to be plane waves in this coordinate. Write the solution as

$$\psi(x, y) = \frac{e^{ik_x x}}{\sqrt{L_x}} f(y) \tag{11.278}$$

$$H\psi(x, y) = \frac{e^{ik_x x}}{\sqrt{L_x}}\frac{1}{2m}\left[\left(\hbar k_x + \frac{eBy}{c}\right)^2 + p_y^2\right] f(y) \tag{11.279}$$

The effective Hamiltonian operating on the y variable is a harmonic oscillator. This observation is made clearer by defining some constants, where the electron charge is ($e<0$)

$$\omega_c = \frac{|e|B}{mc} \tag{11.280}$$

$$l^2 = \frac{\hbar}{m\omega_c} = \frac{\hbar c}{|e|B} \tag{11.281}$$

$$y_0 = -\frac{\hbar c k_x}{eB} = k_x l^2 \tag{11.282}$$

$$H\psi(x,y) = \frac{e^{ik_x x}}{\sqrt{L_x}}\left[\frac{p_y^2}{2m} + \frac{m\omega_c^2}{2}(y-y_0)^2\right]f(y) \tag{11.283}$$

The last equation provides the effective Hamiltonian in the y direction in the explicit form of the harmonic oscillator equation. Its eigenstates are harmonic oscillator functions $\phi_n(y-y_0)$ with eigenvalue $E_n = \hbar\omega_c(n+\frac{1}{2})$. The solution to the Hamiltonian is

$$\psi(x,y) = \frac{e^{ik_x x}}{\sqrt{L_x}}\phi_n(y-y_0) \tag{11.284}$$

$$E_{n,m_s} = \hbar\omega_c\left[n+\tfrac{1}{2}\right] + m_s\mu B \tag{11.285}$$

The magnetic length scale is l. The last term is for the Pauli interaction, where $m_s = \pm\frac{1}{2}$. The Landau levels are states with different values of n. The lowest Landau level has $n=0$. It is interesting that the eigenvalue does not depend upon the quantum number k_x. The state has a high degeneracy, in that many different values of k_x have the same eigenvalue. The classical picture has the electron going in a circular orbit. The center of the orbit can be located any where in the system, without altering the energy. For the quantum system, the value of k_x relates to the location of the center of the harmonic oscillator.

A system of one electron will have a ground state with the electron in the lowest Landau level ($n=0$). The interesting question is how these states are occupied for a set of N_e electrons. The question can be rephrased: In a system of area $A = L_x L_y$, how many electrons can be put into the lowest Landau level? The constant y_0 is the center of the cyclotron orbit. The orbit center must be in the area, which provides the constraint

$$0 \le y_0 = k_x l^2 \le L_y \tag{11.286}$$

$$k_x = \frac{2\pi n_x}{L_x} \le \frac{L_y}{l^2} \tag{11.287}$$

$$n_x \le \frac{L_x L_y}{2\pi l^2} = \frac{m\omega_c}{2\pi\hbar}A \equiv n_{\max} \tag{11.288}$$

Although the energy does not depend upon the value of k_x, this quantum number has a restricted range of value. The requirement of keeping the orbit center in the area constrains $n_x \le n_{\max}$.

Define $\rho(E)$ as the density of states per unit area. This state density is

$$\rho(E) = \frac{1}{L_x L_y}\sum_{n,n_x,m_s}\delta(E - E_{n,m_s}) \tag{11.289}$$

where E_{n,m_s} is the eigenvalue in (11.285). The summation over n_x just gives $n_{\max}$. Then the factor of area cancels

$$\rho(E) = \frac{m\omega_c}{h} \sum_{n,m_s} \delta(E - E_{n,m_s}) \tag{11.290}$$

$$n_0 = \int dE \rho(E) n_F(E) = \frac{m\omega_c}{h} \sum_{n,m_s} n_F(E_{n,m_s}) \tag{11.291}$$

$$\frac{m\omega_c}{h} = \frac{1}{2\pi l^2} = \frac{e}{hc}B = \frac{B}{\phi_0} \tag{11.292}$$

$$n_0 = \frac{\nu}{2\pi l^2} \tag{11.293}$$

$$\nu = \sum_{n,m_s} n_F(E_{n,m_s}) \tag{11.294}$$

The density of electrons n_0 (#/meter^2) can be expressed as a prefactor times the number of occupied Landau levels, where the Fermi–Dirac occupation number is $n_F(E)$. Note that the prefactor can be written as $m\omega_c/h = B/\phi_0$ where the quantum of flux is $\phi_0 = hc/e$. The factor of ν is called the *filling factor* and determines the number of occupied Landau levels.

There are two length scales in the problem. One of them is the magnetic length l. It determines the radius of the harmonic oscillator orbits. The other is the average separation between electrons, which is called a. Highly correlated liquids usually have short range order in the form of a triangular lattice with each electron having six neighbors at a distance a. As an example, examine a woodpile stacked with equally sized logs. The area per electron in the plane triangular lattice is $a^2\sqrt{3}/2$. Setting this formula equal to the inverse of the electron density gives that $a^2 = 2/n_0\sqrt{3}$. At what value of the magnetic field does the separation between electrons a become larger than the orbital diameter? The harmonic oscillator wave function for the lowest Landau level is $\exp[-(y - y_0)^2/2l^2]$. The exponent is -1 at $\Delta y = \sqrt{2}l$. Taking this to be the radius of the orbit, the separation between electrons is greater than the orbit diameter when

$$a > 2\sqrt{2}l \tag{11.295}$$

$$\frac{2}{n_0\sqrt{3}} = a^2 > 8l^2 = \frac{4\phi_0}{\pi B} \tag{11.296}$$

$$\frac{\pi}{2\sqrt{3}} > \frac{n_0\phi_0}{B} = \sum_{m_s,n} n_F(E_n) = \nu \tag{11.297}$$

The number on the left is about 0.907. The summation over the occupied levels, on the right in the above equation, must be a less than 0.9 before the Landau levels are truly separated in space. These values are reached in the FQHE for fractions such as $\nu = \frac{1}{3}$ or $\frac{2}{5}$.

11.3.2. Classical Hall Effect

The classical Hall effect must be understood before its quantum version can be appreciated. The measurement geometry is shown in Fig. 11.16. The shaded area is the experimental sample. A current I is sent through the material, along the direction labeled "x". The voltage change V_x is measured between the two leads on top. This configuration is a standard four-probe measurement. The voltage contacts are not put at the end of the sample, in order to avoid the influence of nonohmic contacts where the current enters and leaves. Another pair of

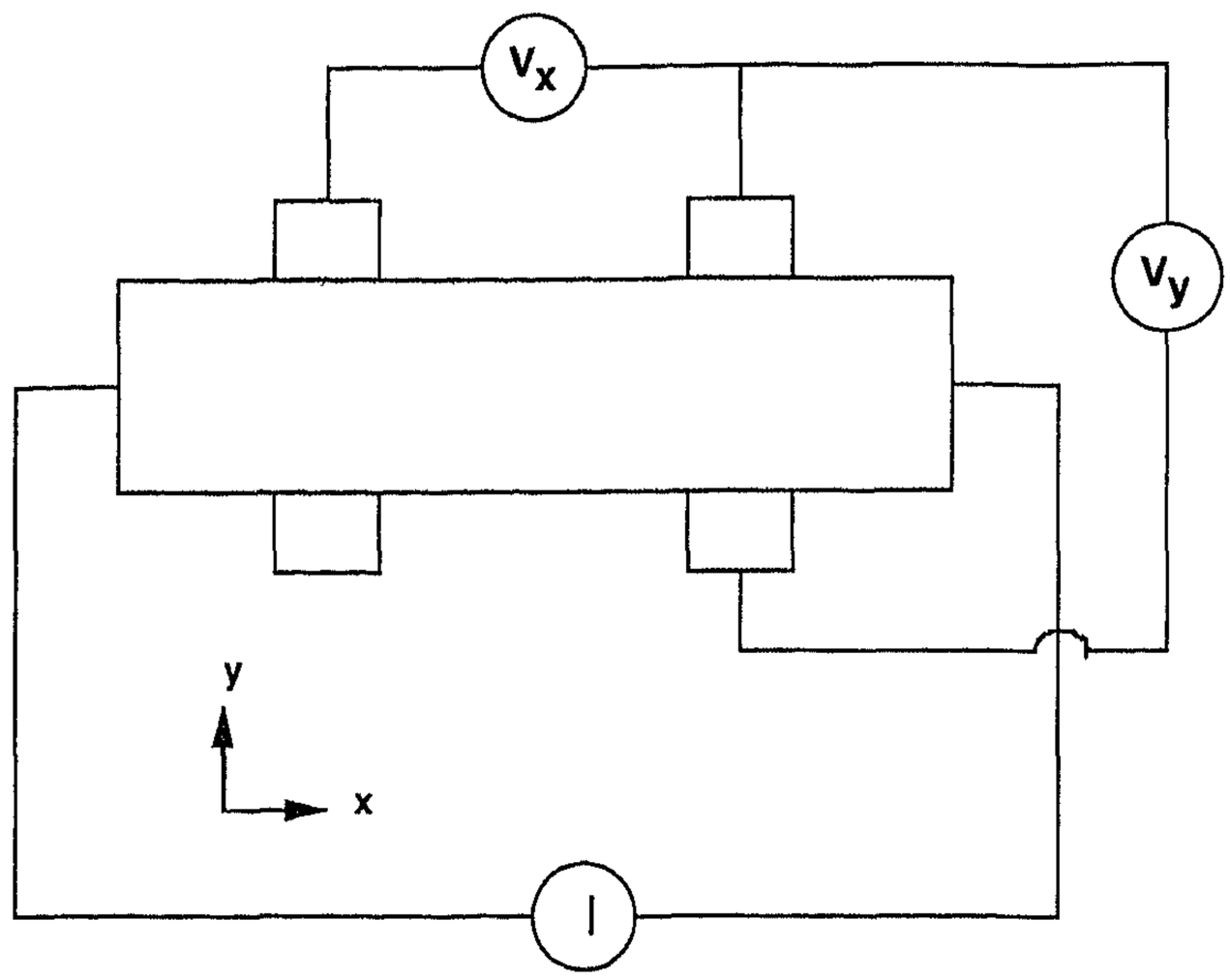


FIGURE 11.16 Experimental geometry for the measurement of the Hall effect. The magnetic field is perpendicular to the plane.

voltage contacts measure the voltage V_y across the sample. This voltage is zero in the absence of a magnetic field.

The Hall effect occurs when there is a magnetic field B perpendicular to the plane of the sample. Then a nonzero value is found for the voltage V_y, which is proportional to both B and I. Since the current I is along the x direction, define two components of resistance

$$R_{xx} = \frac{V_x}{I} \tag{11.298}$$

$$R_{yx} = \frac{V_y}{I} \tag{11.299}$$

For most materials, the magnetic field hardly changes R_{xx}. The main experimental observation is the nonzero value of R_{yx}.

The physics is simple when considering the motion of the electrons as classical. Assume the particles obey Newtonian mechanics. In two dimensions, the magnetic field creates a Lorentz force

$$m\dot{\mathbf{v}} = e\left[\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}\right] - \frac{m\mathbf{v}}{\tau} \tag{11.300}$$

$$m\dot{v}_x = e\left[E_x + \frac{v_y B}{c}\right] - \frac{mv_x}{\tau} \tag{11.301}$$

$$m\dot{v}_y = e\left[E_y - \frac{v_x B}{c}\right] - \frac{mv_y}{\tau} \tag{11.302}$$

where $\mathbf{E}$ is the electric field and τ is the relaxation time for scattering. Consider that the experiment is steady state, so that the time derivatives are zero. The currents densities are given by $J_x = en_0 v_x$, $J_y = en_0 v_y$. There is no current in the y direction since the devices to measure

voltage have very high impedance. Setting $v_y = 0$ in the last equation gives the identity

$$E_y = \frac{v_x B}{c} = \frac{J_x B}{e n_0 c} = \frac{V_y}{L_y} \tag{11.303}$$

$$V_y = J_x B \frac{L_y}{e n_0 c} \tag{11.304}$$

$$v_x = \frac{e\tau}{m} E_x, \qquad J_x = \sigma E_x, \qquad \sigma = \frac{n_0 e^2 \tau}{m} \tag{11.305}$$

The Lorentz force tends to drive the electrons in the perpendicular direction. The flow of electrons along the x direction is diverted to the y direction. Since no current can flow in the y direction, then a voltage V_y must be be built up to force the electrons back along the x direction. The voltage is due to an accumulation of excess charge along the top of the device.

The quantity J_x is the current density. In a three-dimenensional system it has the units of amperes per area, so that $I_x = J_x A_0$ where A_0 is the cross sectional area of the sample. In the present case $A_0 = L_y t$, where t is the thickness of the sample. One then gets in 3D that

$$V_y = \frac{I_x B}{n_0 e c t} \tag{11.306}$$

The Hall effect is very useful in the laboratory for measuring the density n_0 of current carrying particles in the system. It is a relatively easy measurement if a sample of material is available which is several millimeters on each side. The Hall measurement gives the density of carriers in semiconductors and metals. The result is only simple to interpret if all carriers are in the same energy band of the crystal.

In two dimensions the Hall effect is even easier to derive. The particle density n_0 has the units of number of particles per area. The current density J_x has the units of amperes per unit width. Therefore $I_x = J_x L_y$ and the expression for the Hall voltage in 2D is

$$V_y = \frac{I_x B}{n_0 e c} \tag{11.307}$$

$$R_{yx} = \frac{B}{n_0 e c} \tag{11.308}$$

This equation does not depend upon any dimension of the sample. Samples of differing shapes and size give the same measurement.

The classical Hall effect has two characteristics:

- The longitudinal resistance R_{xx} is independent of magnetic field.
- The transverse resistance R_{yx} is proportional to magnetic field.

These characteristics change in the Quantum Hall Effect.

11.3.3. Quantum Hall Effect

A quasi-two-dimensional electron gas can be created in semiconductor microstructures. There are several ways to accomplish this feat, and most of them have shown the QHE. One system is the channel of a field-effect transistor. The gate voltage is used to confine the electrons in the direction perpendicular to the plane, where the barrier oxide traps them along the interface. The electrons are free to move in the plane, and their motion is largely two dimensional. In order to observe the QHE (quantum Hall effect), the semiconductor sample must have exceptionally high purity and a large value for the mobility. Experiments originally showed the QHE in silicon MOSFETs. Another kind of semiconductor microstructure is a

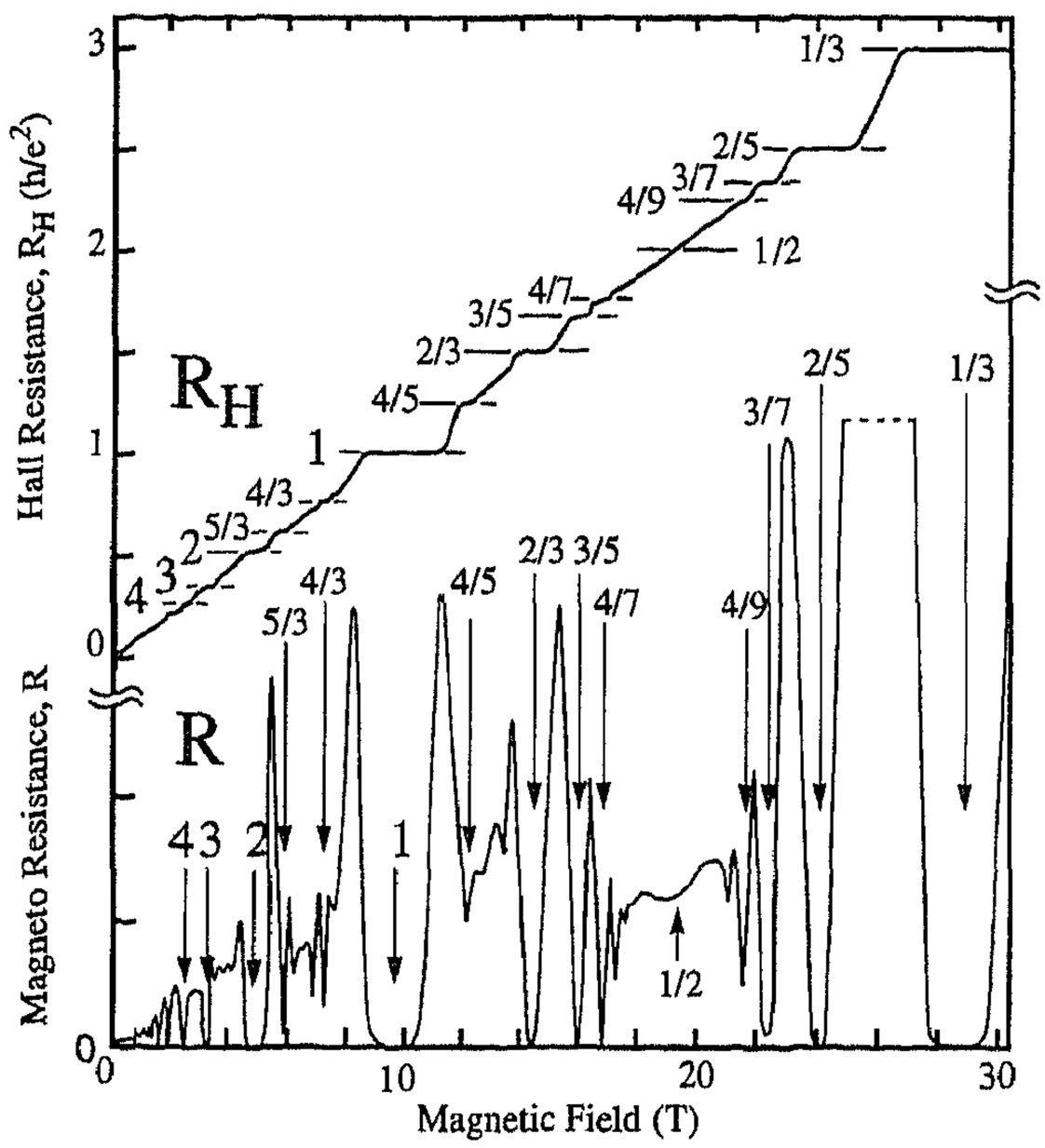


FIGURE 11.17 Quantum Hall effect and fractional quantum Hall Effect in GaAs. (Störmer, 1999 used with permission).

quantum well. It is formed by growing n-layers of one semiconductor, such as GaAs, between another semiconductor such as AlAs. Since AlAs has a larger energy gap, electrons in the conduction band of GaAs are confined in the direction perpendicular to the layers. They move readily along the layer, so the motion is quasi-two-dimensional. Since GaAs and AlAs have the same lattice constant, they can be alloyed continuously. The barrier layers are often $Al_{1-x}Ga_xAs$. Most measurements have been done in single quantum wells of GaAs, although the phenomena has been observed in a few other semiconductors.

A typical result for the QHE and FQHE is shown in Fig. 11.17, which shows the resistivity as a function of the magnetic field B. The two resistances have the following features:

- The transvere resistance $R_{yx} \equiv R_H$ develops plateaux at values of $R_H = R_0/i$, where $R_0 = h/e^2 \approx 25812.8\Omega$, and i is an integer or simple fraction. The QHE is when i is integer, the FQHE is when i is a fraction.
- The longitudinal resistance $R_{xx} \equiv R \to 0$ during these plateaux.

The quantity R_0 is a combination of fundamental constants. It has become the world standard for maintaining electrical resistance. Since the resistance steps appear at the same values, regardless of the shape of the sample, any laboratory in the world will find the steps at the same numerical values.

In the earlier discussion of Landau levels, an expression was derived for the density of electrons in the plane. Using that expression (11.291) gives

$$\frac{1}{R_{yx}} = \frac{n_0 ec}{B} = \frac{e^2}{h}\nu \tag{11.309}$$

$$\nu = \sum_{n,m_s} n_F[\beta(E_{n,m_s} - \mu)] \tag{11.310}$$

$$R_{yx} = \frac{R_0}{\nu} \tag{11.311}$$

The QHE has a simple explanation if the function ν has integer values. The FQHE requires that the filling factor ν be a simple fraction.

The key to understanding the QHE is to understand the behavior of the function ν as the magnetic field in increased. The important parameter is the chemical potential μ. How does it change with magnetic field? There are three different models for the behavior of the chemical potential in these systems.

11.3.3.1. Fixed Density

The first model assumes that the density of electrons n is fixed in the sample. When the magnetic field B is varied, the chemical potential must also be varied so that the density n_0 is unchanged. When considering a function such as (11.291), in order to keep the right-hand side of this equation a constant as B is varied, the chemical potential must go through some gyrations. The product $n_0\phi_0 = B_0$ has the units of magnetic field. For small values of B, $(B \ll B_0)$, the summation over n needs many terms. Then the chemical potential varies with magnetic field in a manner which is periodic in $1/B$. This behavior is well known, and shows up experimentally in de Haas–van Alphen, Shubnikov–deHaas, and other measurements of these oscillations. However, as B increases and becomes $B \geq B_0$ then all electrons are in the lowest Landau level. The requirement of constant density gives the expression for the chemical potential (neglecting the Zeeman term)

$$B_0 \approx \frac{B}{e^{\beta(\hbar\omega_c/2-\mu)} + 1} \tag{11.312}$$

$$\lim_{B \gg B_0} \mu = \tfrac{1}{2}\hbar\omega_c - k_B T \ln\left(\frac{B}{B_0} - 1\right) \tag{11.313}$$

The variation of the chemical potential with magnetic field is an important feature of electrons in large magnetic fields.

This model is applied to most three-dimensional solids. There the number of electrons is fixed by considerations of charge neutrality.

11.3.3.2. Fixed Chemical Potential

The second model for the chemical potential assumes that it does not change with magnetic field. Note that in the experimental measurements, there are at least five metallic wire leads soldered to the sample. These metal contacts form a reservoir of electrons which could fix the chemical potential in the device. The two-dimensional device has only about 10^{11} conduction electrons. That is a very small number compared to the number of electrons in the metallic reservoirs. The change in energy of these small number of electrons has a negligible contribution to the energy of the system. The chemical potential is thereby fixed. The assumption of a fixed chemical potential certainly applies to MOSFET ("Metal-Oxide-Semiconductor-Field Effect Transistor") devices. The usefulness of the gate voltage is that one can use it to modify the density of electrons in the conducting channel. The electrons in the channel have a negligible effect on the chemical potential, which is determined by the source and drain.

In the case that μ is a constant, then the function f will have plateaux at low temperature when plotted as a function of magnetic field.

Figure 11.18 shows a plot of the series

$$f(\mu) = \sum_n n_F[\beta(E_n - \mu)], \qquad E_n = (n + \tfrac{1}{2})\hbar\omega_c \tag{11.314}$$

which is the number of electrons in Landau levels, of a single spin component, as a function of chemical potential. When $\hbar\omega_c \gg k_B T$ there are plateaux in the occupation number f. Figure 11.18(b) shows the value of n_0 as the magnetic field is varied. Note that $n_0 \to 0$ as the lowest Landau level exceeds the chemical potential ($\hbar\omega_c > \mu$). This prediction is in direct contradiction to the experimental data, which shows that n_0 is relatively constant at high magnetic field. Note that the resistance R_{yx} tends linearly upward in Fig. 11.17, except for the plateaux. The linear behavior is only possible when n_0 is relatively constant.

The model with constant chemical potential explains well the QHE (Yennie, 1987). The experimental plateaux are very well fit by this model. The plateaux appear exactly as in Fig. 11.18, which plots f, while the experiments measure $1/f$. However, this model fails to explain the FQHE, and fails to explain the data at large magnetic field which shows that n_0 is relatively constant.

11.3.3.3. Impurity Dominated

The third model for the chemical potential invokes localization caused by static impurities (Prange, 1987). All semiconductors have impurities. They tend to bind the Landau levels near potential mimima in the plane. These minima are due to impurities in the underlying substrate. In one dimension, all states are localized by any impurity scattering. In the QHE, the system is two dimension. However, one quantum number becomes localized into a harmonic oscillator state. The second quantum number (k_x) defines a type of one-dimensional motion. The particles behave, in some respects, as if they were in one dimension. However, their bound states are not given by a Hamiltonian in one dimension, since there is no kinetic energy term associated with the motion in the x direction.

The argument in favor of localization caused by impurities, as controlling the chemical potential, is that as the experimental samples are made with higher purity, the plateaux become narrower. Inpurities are essential for pinning the value of chemical potential, and thereby observing the FQHE.

The fractional quantum Hall effect (FQHE) is the appearance of additional plateaux at noninteger value of the function ν. Tsui *et al.* (1982) discovered that increasing the magnetic field, beyond the plateaux for $\nu = 1$, produced additional plateaux which appear at fraction occupation such as $\frac{1}{3}$ or $\frac{2}{5}$. The experimental technique continues to improve, with samples of higher purity, at larger magnetic fields, and at lower temperature. Figure 11.17 shows one of their recent experimental result on a sample of very high purity. Numerous fractional plateaux are apparent.

The plateaux at fractional filling are due to electron correlation. The theory of this phenomena shares some features with liquid helium. It is impossible to describe the correlated states with any kind of perturbation theory. The theory has no small parameter on which to base a perturbation expansion. Instead, progress has occurred by writing down ad hoc wave functions, and seeing what they give. The best wave function is described in the next section.

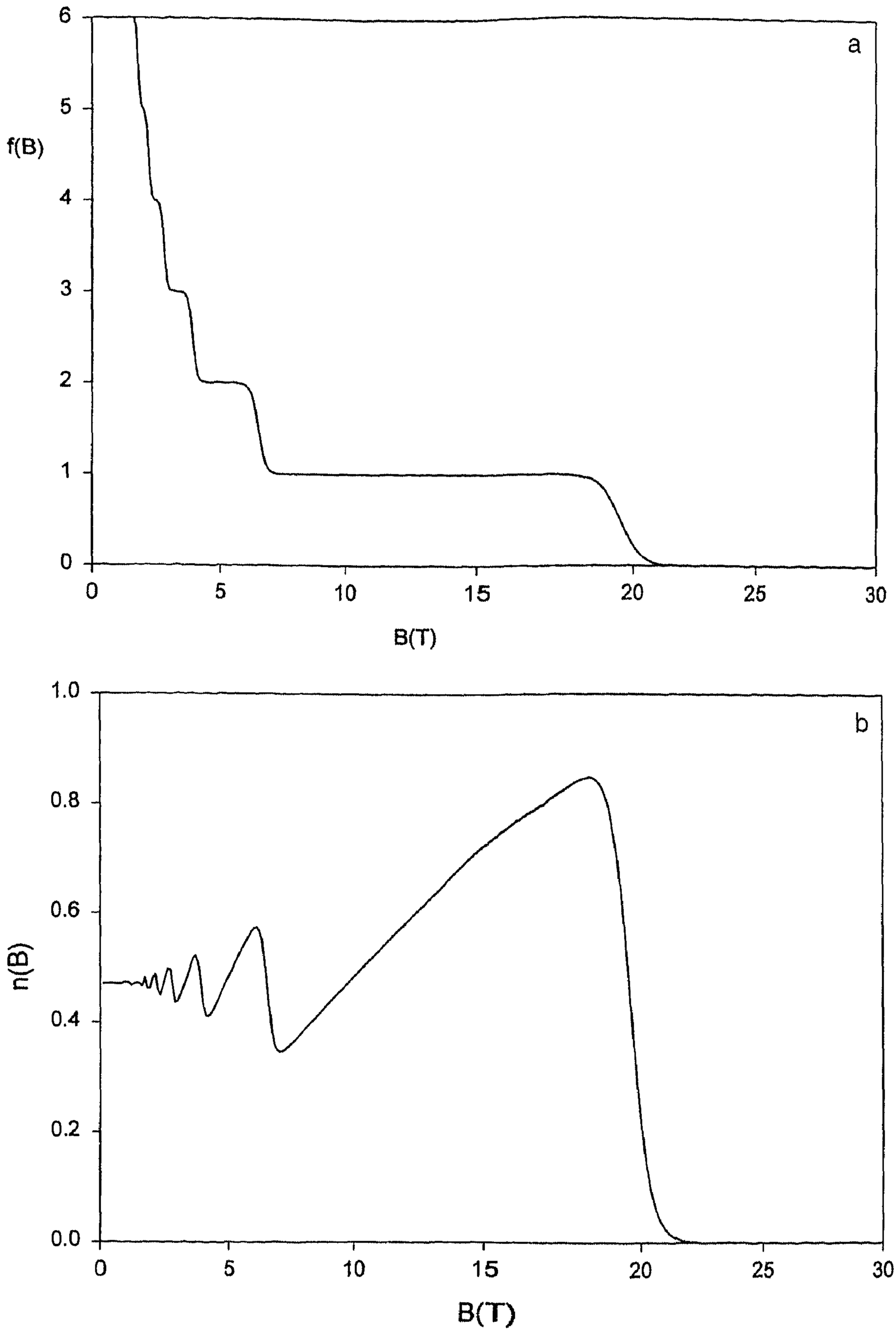


FIGURE 11.18 (a) Number of electrons in Landau levels as a function of magnetic field for a fixed chemical potential. (b) The density of electrons vs. magnetic field, in units of 10^{16} m^{-2}. Parameters are $T = 0.1$ K, $m^* = 0.07$, $n_0 = 0.47 \times 10^{12}$ m^{-2}, and $\mu = 30\ \mu E_{Ry}$.

11.3.4. Laughlin Wave Function

Laughlin proposed a wave function for the FQHE which has become the standard basis for all theories. It is discussed in this section. As with all wave functions in this chapter, it is written down by intuition, and contains correlation. The correlated wave function is not derived by any form of perturbation theory. It is just written down, and its properties are studied.

The first step in the derivation is to solve the Hamiltonian for the symmetric gauge (11.274). For a single electron, the Hamiltonian in this gauge is

$$H = \frac{1}{2m}\left[p_x + \frac{eB}{2c}y\right]^2 + \frac{1}{2m}\left[p_y - \frac{eB}{2c}x\right]^2 \tag{11.315}$$

The space variables are put in dimensionless form using $x' = x/l, y' = y/l$ where the magnetic length is $l^2 = \hbar c/eB$. After some algebra, the Hamiltonian is rewritten as (dropping the primes)

$$H = \frac{\hbar\omega_c}{2}\left[\left(\frac{\partial}{i\partial x} + \frac{y}{2}\right)^2 + \left(\frac{\partial}{i\partial y} - \frac{x}{2}\right)^2\right] \tag{11.316}$$

$$= \frac{\hbar\omega_c}{2}\left[-\nabla^2 + \frac{1}{4}\rho^2 - L_z\right] \tag{11.317}$$

$$L_z = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) = -i\frac{\partial}{\partial\phi} \tag{11.318}$$

where $\tan(\phi) = y/x$ and $\rho^2 = x^2 + y^2$ in polar coordinates. Consider the eigenfunction $\psi_m(\rho, \phi)$

$$\psi_m(\rho, \phi) = N_m \rho^m e^{im\phi} e^{-\rho^2/4} \tag{11.319}$$

$$H\psi_m = \frac{\hbar\omega_c}{2}\psi_m \tag{11.320}$$

$$N_m = \frac{1}{\sqrt{2\pi m! 2^m}} \tag{11.321}$$

$$\delta_{mm'} = \int_0^\infty \rho d\rho \int_0^{2\pi} d\phi \psi_m^* \psi_{m'} \tag{11.322}$$

The eigenfunctions ψ_m have an eigenvalue of $\hbar\omega_c/2$ which indicates that the electron is in the lowest Landau level ($n = 0$). The quantum number m indicates angular momentum, since $L_z\psi_m = m\psi$. The value of angular momentum m does not alter the eigenvalue. In the prior section the silent quantum number was k_x, which also did not affect the eigenvalue. In the symmetric gauge the silent quantum number is m. The last expression shows that these functions are mutually orthogonal, because the integral over $d\phi$ vanishes unless $m = m'$.

In the FQHE, the high value of magnetic field at low temperature ensures that $\hbar\omega_c \gg k_B T$. All electrons are in the lowest Landau level for the case that the fractional filling $\nu < 1$. Then the eigenfunction ψ_m is the basis for all electrons in the ground state. An

interesting question is to ask where are the electrons? Use Stirling's approximation to the factorial to write the electron density for large m as

$$\rho_m(\rho) = |\psi_m|^2 = \frac{1}{2\pi m! 2^m}\rho^{2m}e^{-\rho^2/2} \tag{11.323}$$

$$\approx \frac{1}{2\pi\sqrt{2\pi m}}\exp\left[-\frac{\rho^2}{2} + m\ln(\rho^2) + m - m\ln(2m)\right] \tag{11.324}$$

$$\approx \frac{1}{2\pi\sqrt{2\pi m}}\exp\left[-\tfrac{1}{2}(\rho^2 - 2m) + m\ln\left(\frac{\rho^2}{2m}\right)\right] \tag{11.325}$$

$$\approx \frac{1}{2\pi\sqrt{2\pi m}}\exp\left[-\tfrac{1}{2}(\rho^2 - 2m) + m\ln\left(1 + \frac{\rho^2 - 2m}{2m}\right)\right]$$

$$\approx \frac{1}{2\pi\sqrt{2\pi m}}\exp\left[-\frac{1}{8m}(\rho^2 - 2m)^2\right] \tag{11.326}$$

For large value of angular momentum m, the peak in the electron density occurs at $\rho \sim \sqrt{2m}$. The electrons are in a circle, where $m = 0$ is the orbit at the center of the circle, and higher values of m are further from the center.

The above discussion is for a single electron. The next step is to construct a many-electron wave function. In this case it is useful to describe the position vector (x, y) by the complex number $z = x + iy = \rho e^{i\phi}$. The basis set is now written as

$$\psi_m(z) = N_m z^m \exp(-|z|^2/4) \tag{11.327}$$

In the Hartree–Fock approximation, the wave function for a many-electron system is a Slater determinant. All electrons are assumed to have the same spin direction since they are aligned by the magnetic field. If they have spin up, the eigenfunction is α_j for the j-th electron. Each electron must be in a different state, so that they occupy the angular momentum states $m = 0, 1, 2, \ldots, N-1$. Then the Hartree–Fock wave function for N electrons is

$$\Psi_N(z_1, z_2, \ldots, z_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_0(z_1) & \psi_1(z_1) & \psi_2(z_1) & \cdots & \psi_{N-1}(z_1) \\ \psi_0(z_2) & \psi_1(z_2) & \psi_2(z_2) & \cdots & \psi_{N-1}(z_2) \\ \psi_0(z_3) & \psi_1(z_3) & \psi_2(z_3) & \cdots & \psi_{N-1}(z_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_0(z_N) & \psi_1(z_N) & \psi_2(z_N) & \cdots & \psi_{N-1}(z_N) \end{vmatrix}$$

$$\times\, \alpha_1 \cdots \alpha_N \tag{11.328}$$

$$= \frac{1}{\sqrt{N!}}[\Pi_k N_k] D_N \alpha_1 \cdots \alpha_N \exp\left[-\frac{1}{4}\sum_{j=1}^{N}|z_j|^2\right] \tag{11.329}$$

$$D_N = \begin{vmatrix} 1 & z_1 & z_1^2 & \cdots & z_1^{N-1} \\ 1 & z_2 & z_2^2 & \cdots & z_2^{N-1} \\ 1 & z_3 & z_3^2 & \cdots & z_3^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_N & z_N^2 & \cdots & z_N^{N-1} \end{vmatrix} \tag{11.}$$

The determinant is due to Vandermonde, and is exactly equal to the product of the difference for all pairs of coordinates

$$\Psi_N(z_1, z_2, \ldots, z_N) = \frac{1}{\sqrt{N!}} \Pi_k N_k \Pi_{j>i}(z_j - z_i)\alpha_1 \cdots \alpha_N \exp\left[-\frac{1}{4}\sum_{j=1}^{N} |z_j|^2\right] \tag{11.331}$$

The first product is the normalization coefficients. The second product contains $(z_j - z_i)$ for all pairs of electrons. This factor ensures the many-electron wave function is anti-symmetric under exchange of any two electrons, as is required for fermions.

The wave function Ψ_N can be used to calculate some properties of the many-electron system. Squaring it, and then integrating over all but one coordinate, gives the average density of one electron. Using the theorems in Sec. 5.1.9, and the fact that the eigenfunctions are orthogonal, gives the result

$$\rho_1(\rho) = \frac{1}{N}\sum_{m=0}^{N-1} |\psi_m|^2 = \frac{e^{-\rho^2/2}}{2\pi N}\sum_m \frac{\rho^{2m}}{2^m m!} \tag{11.332}$$

Again the notation ρ_1 means the density of one particle in the many-particle system. For small values of ρ, the summation gives $\exp(\rho^2/2)$ which cancels the other exponent, and $\rho_1 \approx 1/(2\pi N)$. The density of N such electrons, after restoring the factors of length l, are

$$n_0 = N\rho_1 \approx \frac{1}{2\pi l^2} \tag{11.333}$$

Equation (11.293) has almost the same expression except the right-hand side is multiplied by ν. Equating these two expressions gives that $\nu = 1$. The many-particle wave function was chosen to fill up the lowest states of angular momentum. This case corresponds to the maximum packing of electrons in the lowest Landau level, which is a filling factor $\nu = 1$.

The pair distribution function $g(r)$ is found from $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$, which is obtained by integrating over all but two coordinates. Since the functions are orthogonal, ρ_2 is found from all pairs of two-particle wave functions

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{N(N-1)}\sum_{ij} |\psi_i(\mathbf{r}_1)\psi_j(\mathbf{r}_2) - \psi_i(\mathbf{r}_2)\psi_j(\mathbf{r}_1)|^2 \tag{11.334}$$

$$= \frac{e^{-(\rho_1^2+\rho_2^2)/2}}{2\pi N(N-1)}\sum_{ij}^{N-1} \frac{1}{2^{i+j} i! j!}\left(\rho_1^{2i}\rho_2^{2j} + \rho_1^{2j}\rho_2^{2i} - 2(\rho_1\rho_2)^{i+j}\cos[(i-j)(\phi_1 - \phi_2)]\right) \tag{11.335}$$

For small values of $\rho_{1,2}$ the series converge to an exponential. In this case the pair distribution function is

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\pi N(N-1)}\left[1 - \exp\left(-\tfrac{1}{2}[\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2\cos(\phi_1 - \phi_2)]\right)\right]$$

$$g(\mathbf{r}_1 - \mathbf{r}_2) = 1 - \exp\left[-\frac{1}{2l^2}(\mathbf{r}_1 - \mathbf{r}_2)^2\right] \tag{11.336}$$

$$S(k) = 1 + n_0\int d^2r[g(r) - 1]e^{i\mathbf{k}\cdot\mathbf{r}} = 1 - \exp[-k^2 l^2/2] \tag{11.337}$$

Note this pair distribution function is for two particles in a system of $N \gg 1$ particles. A different result is obtained for only a few particles. One can also derive the three- and four-particle distribution functions (Jancovici, 1981). The static structure factor $S(k)$ is also given.

Note that it goes to zero as $O(k^2)$ rather than as $O(k)$ as in a fluid of neutral particles. These formulas apply only for the case of $\nu = 1$.

In Chap. 5 the exchange energy per electron of the electron gas is provided by the formula

$$E_x = -\frac{n_0 e^2}{2}\int \frac{d^2 r}{r}[1 - g(r)] \tag{11.338}$$

where it has been adopted to two dimensions by changing $d^3r \to d^2r$. Using the above expression for $g(r)$, and $n_0 = 1/(2\pi l^2)$ for $\nu = 1$, the exchange energy is ($r = \rho/l$)

$$E_x = -\frac{e^2}{2l}\int_0^\infty dr e^{-r^2/2} = -\frac{e^2}{2l}\sqrt{\frac{\pi}{2}} \tag{11.339}$$

This formula is valid for filling factor $\nu = 1$. For fractional filling, the exchange energy in the Hartre–Fock approximation is written as $E_x = -C_\nu e^2/l$. Using the Laughlin wave function, described below, gives the following values for the dimensionless constant:

$1/\nu$	C_ν
1/1	0.627
1/3	0.416
1/5	0.334

The Laughlin wave function for the FQHE was written down by intuition starting from Eq. (11.331). It is

$$\Psi(z_1, z_2, \ldots, z_N) = N_{pN}\Pi_{j>i}(z_j - z_i)^p \alpha_1 \cdots \alpha_N \exp\left[-\tfrac{1}{4}\sum_{j=1}^{N}|z_j|^2\right] \tag{11.340}$$

where the normalization constant is N_{pN}. The factor of $(z_j - z_i)^p$ is the difference between the coordinates raised to a power of p. The fermion wave function has to change sign under the exchange of any two coordinates. This fundamental antisymmetrization is achieved as long as p is any odd integer. The case with $p = 1$ is the integer QHE with $\nu = 1$ in (11.331). Other values such as $p = 3, 5, 7 \cdots$ provide an accurate wave function for the FQHE. The Laughlin wave function cannot be derived from a single Slater determinent and is not a Hartree–Fock wave function. It contains correlation and is a highly correlated wave function.

The exponent p is related to the filling factor $p = 1/\nu$. The choice $p = 3$ gives a wave function for the fractional state $\nu = \frac{1}{3}$. The proof of this assertion starts by examining the density matrix

$$\rho_N(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = |\Psi(z_1, z_2, \ldots, z_N)|^2 \tag{11.341}$$

$$= N_{pN}^2 \Pi_{i>j}|\mathbf{r}_i - \mathbf{r}_j|^{2p} \exp\left[-\frac{1}{2}\sum_j r_j^2\right] \tag{11.342}$$

$$= N_{pN}^2 \exp\left[p\sum_{i>j}\ln|\mathbf{r}_i - \mathbf{r}_j|^2 - \frac{1}{2}\sum_j r_j^2\right] \tag{11.343}$$

The spin eigenstates have been eliminated. Also note that

$$|z_1 - z_2|^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 = (\mathbf{r}_1 - \mathbf{r}_2)^2.$$

The density matrix can be cast into the form for a classical fluid $\rho_N \sim \exp[-\beta V]$ where the potential energy $\beta V \equiv U$ has the form

$$U(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \frac{1}{2}\sum_j r_j^2 - p\sum_{i>j} \ln|\mathbf{r}_i - \mathbf{r}_j|^2 \tag{11.345}$$

The potential has two types of terms. Each electron has a potential $r_i^2/2$ which tends to confine it to the origin. There is also a particle–particle interaction of the form of a logarithm. It is characteristic of the two-dimensional Coulomb gas, where the particles are not points with $V = e^2/r_{ij}$, but instead are charged rods perpendicular to the plane which have $V = e^2 \ln(r_{ij})$. The same potential is found for interacting vortices.

For a system of N electrons, the two-dimensional fluid forms a circle of radius $R \sim \sqrt{2N}$. The electrons near to the center of the circle are in equilibrium, which means that the force on them should be zero. For a uniform circle, the force can only be radial. The radial force on one of them, say $\mathbf{r}_1$ is

$$F_r(r_1) = -\frac{\partial U}{\partial r_1} = -r_1 + p\sum_{j=2}^{N} \frac{2r_1 - 2r_j\cos(\phi_j)}{r_1^2 + r_j^2 - 2r_1 r_j \cos\phi_j} \tag{11.346}$$

where ϕ_j is the angle between $\mathbf{r}_1$ and $\mathbf{r}_j$. The last term is evaluated by changing the summation to an integral, assuming the particles have a uniform density of n_0

$$F_r(r_1) = -r_1 + pn_0 \int d^2 r_j \frac{2r_1 - 2r_j\cos(\phi_j)}{r_1^2 + r_j^2 - 2r_1 r_j \cos\phi_j} \tag{11.347}$$

$$= -r_1 + \frac{pn_0}{r_1}\int d^2 r_j \left[1 + \frac{r_1^2 - r_j^2}{r_1^2 + r_j^2 - 2r_1 r_j \cos\phi_j}\right] \tag{11.348}$$

The angular integral is done using

$$\int_0^{2\pi} \frac{d\phi}{A + B\cos\phi} = \frac{2\pi}{\sqrt{A^2 - B^2}}\Theta(A^2 - B^2) \tag{11.349}$$

$$F_r(r_1) = -r_1 + \frac{2\pi p n_0}{r_1}\int r_j dr_j \left[1 + \frac{r_1^2 - r_j^2}{|r_1^2 - r_j^2|}\right] \tag{11.350}$$

The bracket in the integrand is zero if $r_j > r_1$ and is two if $r_j < r_1$. In fact, the above expression is Gauss's law for the two-dimensional Coulomb gas. The force comes only from the charge inside of the circle of radius r_1. The remaining integral is elementary,

$$F_r(r_1) = -r_1 + \frac{4\pi p n_0}{r_1}\int_0^{r_1} r_j dr_j = -r_1[1 - 2\pi n_0 p] \tag{11.351}$$

The density $n_0 = \nu/2\pi$ in units where all dimensions are scaled by the magnetic length l. Then the factor in brackets is $[1 - p\nu]$. The requirement that the force on particle 1 be zero sets $p = 1/\nu$. Since p is an odd integer, then the Laughlin wave function describes the fractional fillings of $\nu = 1/p$ such as $\frac{1}{3}, \frac{1}{5}, \frac{1}{7}$, etc. Numerous numerical tests have shown that the wave function describes these states extremely well.

The wave function for $p = 1$ is the Slater determinant which gives (11.331). By filling up all states $m = 0, 1, 2, \ldots, N-1$, the system has a uniform density given by $n_0 = 1/(2\pi l^2)$. The Laughlin wave function for $\nu = 1/p$ also has a uniform density, on the average.

It is interesting to calculate the current density associated with the Laughlin ground state. First calculate the current from a typical electron, such as z_1, using the standard quantum mechanical expression

$$\mathbf{J}_1 = \frac{e\hbar}{2mi}\{\Psi\nabla_1\Psi^* - \Psi^*\nabla_1\Psi\} \tag{11.352}$$

$$\nabla_1\Psi = \Psi\left\{-\frac{\mathbf{r}_1}{2} + p\sum_{i=2}^{N}\frac{\hat{x}+i\hat{y}}{z_1 - z_i}\right\} \tag{11.353}$$

$$\mathbf{J}_1 = -\frac{pe\hbar}{m}|\Psi|^2\sum_{i=2}^{N}\frac{\hat{x}(y_1-y_i)-\hat{y}(x_1-x_i)}{(\mathbf{r}_1-\mathbf{r}_i)^2} \tag{11.354}$$

The summation over i can be changed to a continuous integral $\sum_i \to n_0\int d^2r$. The factor of $|\Psi|^2$ supplies an additional factor of $1/A$, where A is the area. The current from particle 1 is

$$\mathbf{J}_1 = \frac{epn_0\hbar}{mA}\int d^2r_i\frac{\hat{x}(y_1-y_i)-\hat{y}(x_1-x_i)}{(\mathbf{r}_1-\mathbf{r}_i)^2} \tag{11.355}$$

The integral is zero for an infinite system. The result is nonzero for a finite system consisting of a circle of radius R and area $A = \pi R^2$. The angular integral have the same form as discussed in Eq. (11.347), so that the integral over dr_i is constrained to values $r_i < r_1$. The result is

$$\mathbf{J}_1 = \frac{epn_0\hbar}{2mA}[\hat{x}y_1 - \hat{y}x_1] \tag{11.356}$$

The current is rotational. This result should not be surprising, since the lowest Landau level is composed of eigenstates with many values of angular momentum, which are all rotating in the same direction. Also note that the current obeys the equation of continuity $\nabla\cdot\mathbf{J} = 0$.

11.3.5. Collective Excitations

Several collective excitations have been identified for the QHE and the FQHE. Two are discussed here.

11.3.5.1. Magnetorotons

Girvin *et al.* (1986) used the Bijl–Feyman theory to calculate the frequency dependence of the density fluctuations in the QHE and the FHQE. The method is identical to that used earlier in this chapter for excitations in ^{4}He. However, the derivation for the FQHE has some new wrinkles. These excitations are called *magnetorotons*.

The excitation of wave vector $\mathbf{k}$ is obtained by multiplying the ground state wave function Ψ_g by the density fluctuation $\rho(\mathbf{k})$ which gives an excited state $\Psi_\mathbf{k}$

$$\rho(\mathbf{k}) = \sum_j e^{i\mathbf{k}\cdot\mathbf{r}_j} \tag{11.357}$$

$$\Psi_\mathbf{k} = \frac{1}{\sqrt{NS(\mathbf{k})}}\rho(\mathbf{k})\Psi_g \tag{11.358}$$

$$\omega(\mathbf{k}) = \langle\Psi_\mathbf{k}(H-E_0)\Psi_\mathbf{k}\rangle = \frac{1}{NS(\mathbf{k})}\langle\Psi_g|\rho^\dagger(\mathbf{k})[H,\rho(\mathbf{k})]|\Psi_g\rangle$$

As in the derivation for ^{4}He, the only term which fails to commute with the density operator is the kinetic energy. Repeating that earlier derivation gives the same result

$$\omega(\mathbf{k}) = \frac{\hbar^2 k^2}{2mS(k)} \tag{11.359}$$

This result is quite universal. It applies to any system, and to any fractional filling.

First examine the case $\nu = 1$ where the lowest Landau level is completely filled. The structure factor for that case is given in (11.337).

$$\omega(k) = \frac{\hbar^2}{2m}\frac{k^2}{1 - e^{-k^2 l^2/2}} \tag{11.360}$$

At large values of k the excitations go to the free-particle kinetic energy $\omega(k) = \varepsilon_k$. At small values of k the result is

$$\lim_{k\to 0} \omega(k) = \frac{\hbar^2}{ml^2} = \hbar\omega_c \tag{11.361}$$

The excitation energy is just the cyclotron energy $\hbar\omega_c$. Of course, this makes physical sense: if the lowest Landau level is filled, the excitation is to the next Landau level which is ($n = 1$). This result is also in agreement with a theorem of Kohn (1961) which states that the magnetophonon excitation energy is always $\hbar\omega_c$, regardless of the filling factor, and regardless of the interactions.

Girvin *et al.* (1986) remark on the accuracy of the Bijl–Feynman formula. Here it gives the exact excitation energy, while for ^{4}He it gave a very approximate excitation energy. They remark that the current from this excitation satisfies the equation of continuity ($\nabla \cdot \mathbf{J} = 0$), while it does not when using the correlated basis function for the ground state of ^{4}He. They also remark that the theory assumes a *Single Mode Approximation*. All of the excitation spectra is in the one state. Most fermion systems have a collective mode (e.g., plasmons) plus electron–hole excitations. At long wavelength the collective mode dominates. Then the single mode approximation is an accurate description of the excitation.

Girvin *et al.* predicted the existence of a new low energy collective excitation for the FQHE. The first guess is to use the Laughlin wave function for the ground state Ψ_g for $p = 3, 5$, etc. According to Kohn's theorem, the calculated excitation energy will still be $\hbar\omega_c$. This large excitation energy is obtained because the density operator $\rho(\mathbf{k})$ has contributions from many Landau levels. However, the low lying excitations from the Laughlin state are probably to states which are different configuration of the lowest Landau level. So they used a projection operator P_n which projects out the states for the Landau level n. In their case, they only wanted P_0. Both numerator and denominator are evaluated in this reduced Hilbert space

$$\omega(\mathbf{k}) = \frac{\bar{f}(k)}{\bar{S}(k)} \tag{11.362}$$

$$\bar{\rho}(\mathbf{k}) = P_0 \rho(\mathbf{k}) P_0 \tag{11.363}$$

$$\bar{S}(k) = \frac{1}{N}\langle \Psi_g | \bar{\rho}^\dagger(\mathbf{k})\bar{\rho}(\mathbf{k}) | \Psi_g \rangle \tag{11.364}$$

$$\bar{f}(k) = \frac{1}{N}\langle \Psi_g | \bar{\rho}^\dagger(\mathbf{k})[H, \bar{\rho}(\mathbf{k})] | \Psi_g \rangle \tag{11.365}$$

The evaluation of this expression required numerical work such as Monte Carlo. They were able to prove some important results at long wavelength. They showed that for both the $\nu = \frac{1}{3}$

and $\nu = \frac{1}{5}$ states that $\omega(0) \neq 0$ and the excitation spectrum has a gap. The constant result for $k \to 0$ is not due to plasmon effects. In two dimensions, the plasmon without a magnetic field goes as $\omega^2(q) = n_0 e^2 q/m$ and vanishes at long wavelength. The gap at long wavelength is due to the rigidity of the Laughlin state. For $\nu = \frac{1}{3}$ both $\bar{f}(k)$ and $\bar{S}(k)$ are proportional to k^4 at long wavelength, so that their ratio goes as a constant.

Figure 11.19 shows the numerical predictions for the low energy excitation. In Fig. 11 19(a) are $S(q), \bar{S}(q)$ for the state $\nu = \frac{1}{3}$. In (b) is shown the excitation energy $\omega(k)$ for $\nu = \frac{1}{3}, \frac{1}{5}$. There is a minimum at nonzero wave vector, which is similar to the roton in superfluid ^{4}He. The name *magnetoroton* has been applied to this low energy excitation at nonzero wave vector. At large wave vectors the neutral excitation is a quasiexciton consisting of a bound state of a quasielectron and quasihole.

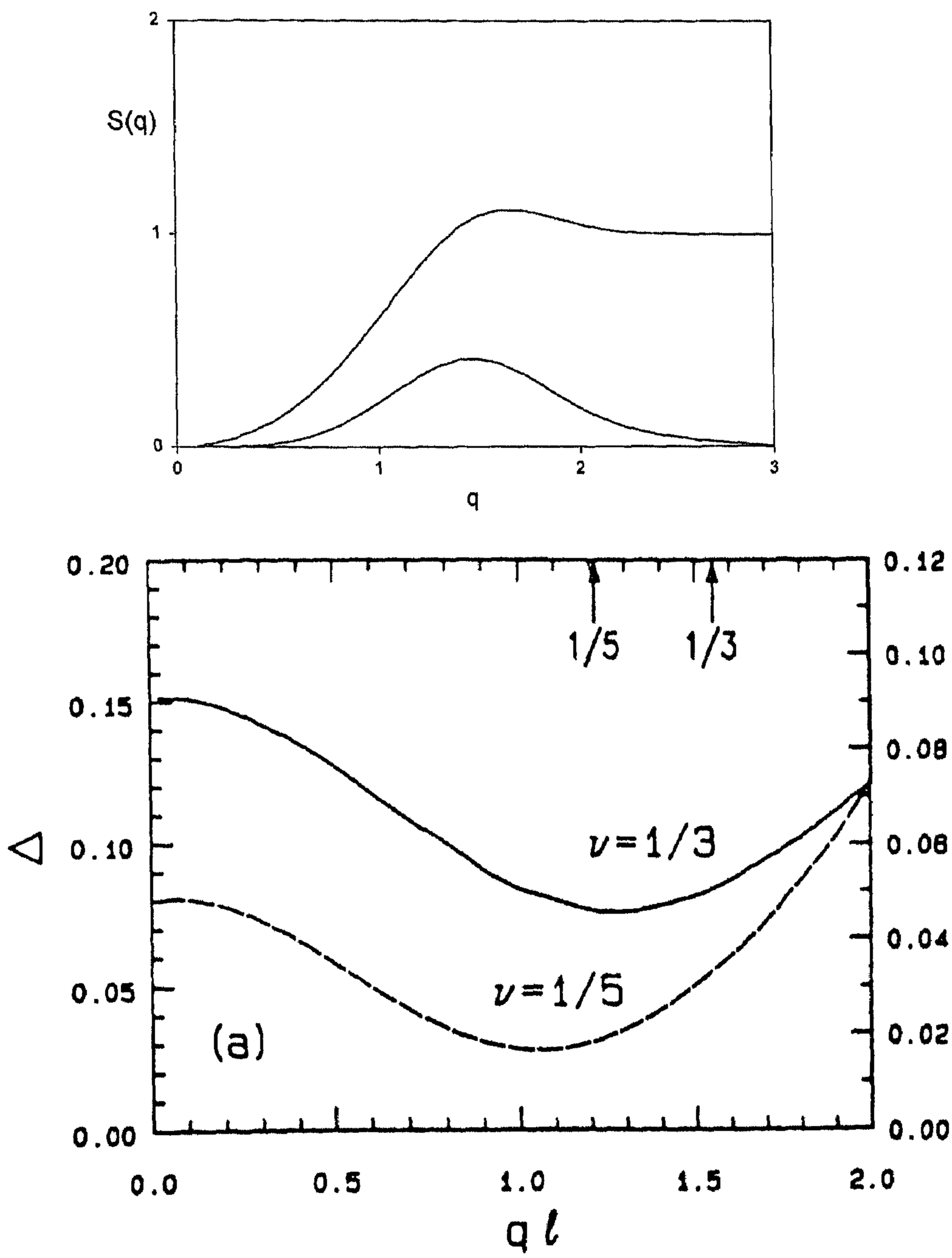


FIGURE 11.19 (a) $\bar{S}(q)$ for states $\nu = \frac{1}{3}$. The top line is $S(q)$ and the lower line is $\bar{S}(q)$. (b) $\omega(k)$ for the states $\nu = \frac{1}{3}, \frac{1}{5}$. From Girvin *et al.* (1986) (used with permission).

11.3.5.2. Quasiholes

Laughlin (1983) considered the excitation created by adding a "quasihole" or "quasielectron". The easiest to understand is the quasihole. It is a place in the fluid where other electrons cannot be located. Denote the Laughlin wave function (11.340) as $\Psi \equiv |p, N\rangle$ for $p = 1/\nu$ and N electrons. The wave function for the fluid with a quasihole at position z_0 is

$$\Psi_N(\{z_i\}; z_0) = W(z_0)|p, N\rangle \tag{11.366}$$

$$W(z_0) = \Pi_i^N (z_i - z_0) \tag{11.367}$$

The factor of $W(z_0)$ ensures that no electron of coordinates z_i will be found at the position z_0. For the case that $p = 1, \nu = 1$ then this orbital factor is equivalent to adding another electron to the system at position z_0. The ground state energy is that of a system with $N + 1$ electrons in the lowest Landau level. The cost in energy from adding the quasihole, in Hartree–Fock, is $\Delta E_h = \hbar\omega_c/2 - \sqrt{\pi/8}\, e^2/l$.

For fractional filling the properties of this excitation must be determined by numerical methods. The probability is found by taking the absolute square of the wave function. The electron density $\rho_1(\mathbf{r}_1, \mathbf{r}_0)$ is found by integrating this probability over all but one space variable. This density must be uniform except near to the point $\mathbf{r}_0$, where it has the form of a pair distribution $g(\mathbf{r} - \mathbf{r}_0)$

$$P(\mathbf{r}_1 \cdots \mathbf{r}_N; \mathbf{r}_0) = |\Psi_N(\{z_i\}; z_0)|^2 \tag{11.368}$$

$$g(\mathbf{r} - \mathbf{r}_0) = \frac{\int P dr_2 \cdots dr_N}{N \int P dr dr_2 dr_3 \cdots dr_N} \tag{11.369}$$

$$q = e \int d^2 r [1 - g(\mathbf{r} - \mathbf{r}_0)] \tag{11.370}$$

The quantity q measures the amount of electron charge which is excluded from the vicinity of the quasihole. It is found that $q = e\nu$. The quasihole has fractional charge.

The evaluation of the above integrals again uses the analogy with interacting fluids. Write P as a Jastrow wave function:

$$P = N(z_0) \exp(-U) \tag{11.371}$$

$$U = -p \sum_{i>j} \ln|\mathbf{r}_i - \mathbf{r}_j|^2 + \frac{1}{2} \sum_j r_j^2 - \sum_i \ln|\mathbf{r}_i - \mathbf{r}_0|^2 \tag{11.372}$$

where $N(z_0)$ is the normalization constant. The fluid analogy has N particles of "charge" $\sqrt{p}$ interacting among themselves with logarthmic forces. An impurity at $\mathbf{r}_0$ of charge $1/\sqrt{p}$ interacts with the other electrons. The resulting pair correlation function is solved using the theory of interacting fluid mixtures, where one component is very dilute. The energy needed to insert the quasihole into the fractional ground state is $E_q = -C_q e^2/l$. Numerical work shows that $C_q = 0.026$ for $\nu = \frac{1}{3}$. For a typical magnetic field of $B = 5$ T, this excitation energy in temperature units is about 6 K. Experiments must be performed at a much lower temperature than this value, in order that quasihole excitations be thermally quenched. A similar energy scale is found for the magnetorotons.

Consider the additional current due to multiplying the Laughlin wave function by a sum of phase factors

$$\Phi = \exp\left[i \sum_i \phi(\mathbf{r}_i) \right] \Psi \tag{11.373}$$

$$\langle \delta \mathbf{J} \rangle = \frac{n_0 \hbar}{m} \nabla \phi(\mathbf{r}) \tag{11.374}$$

The additional current is given by the gradient of the phase factor. If the quasihole is located at $z_0 = 0$ then its factor has the above form

$$W = \Pi_i z = \exp\left[\sum_i \ln(r_i e^{i\theta_i}) \right] \tag{11.375}$$

$$i\phi(\mathbf{r}) = \ln r_i + i\theta \tag{11.376}$$

$$\langle \delta \mathbf{J} \rangle = \frac{n_0 \hbar}{m} \frac{\hat{\theta}}{r} \tag{11.377}$$

The additional current from the quasihole is also rotational. Girvin (1987) points out that the quasihole has the form of a vortex flow.

In many two-dimensional systems the low energy excitations are vortices. The statistical mechanics of the disordering process are dominated by the binding together of vortices of opposite rotational motion. This phenomena was first described by Kosterlitz and Thouless (1973). However, the vortices in the QHE and FQHE are different than those of Kosterlitz and Thouless, and their theory does not actually apply to the present system. These differences are elucidated by Girvin (1987).

The Laughlin wave function explains quite well the states with filling $\nu = 1/p$ where p is an odd integer. The experimentals show many other fractions, and these are not explained by the Laughlin wave function. There are various schemes to explain these states, such as one proposed by Jain (1990). Another interesting feature is that the higher Landau levels show striped phases at very high magnetic field (Lilly *et al.*, 1999).

PROBLEMS

1. Derive the formula for the Bose–Einstein transition temperature T_λ of a gas of bosons in two dimensions.

2. In the Bogoliubov theory of superfluid ^{4}He, derive the terms in the interaction potential which contain three C operators whose wave vectors are not zero. Identify the physics behind each different kind of term.

3. Use CBFs to derive the McMillan result that the average energy per particle is

$$\frac{\langle E \rangle}{N} = \frac{\rho}{2} \int d^3 r g(r) \left[V(r) + \frac{\hbar^2}{2m} \nabla^2 u(r) \right] \tag{11.378}$$

4. Use the result (11.91) for $S(Q, \omega)$ to evaluate the moments ($T = 0$)

$$\int_{-\infty}^{\infty} \frac{d\omega}{4\pi} sgn(\omega) S(Q, \omega) = 2S(Q) \tag{11.379}$$

$$\int_{-\infty}^{\infty} \frac{d\omega}{4\pi} \omega S(Q, \omega) = ? \tag{11.380}$$

$$\int_{-\infty}^{\infty} \frac{d\omega}{4\pi} \omega^3 S(Q, \omega) = ? \tag{11.381}$$

5. Consider the case where the liquid helium is flowing down the pipe with velocity v_s, and the pipe walls are moving with velocity v_N. Derive an expression for the total momentum as a function of ρ_s, ρ_N, v_s, v_N.

6. For the superfluid ^{4}He flowing through a pipe, derive an expression for the total energy to order $O(v^2)$. Show it is not exactly given by the superfluid density ρ_s.

7. Use the paramagnon Hamiltonian (11.208) to evaluate the self-energy of quasiparticles by summing the set of diagrams in (a) Fig. 11.20(a) and (b) Fig. 11.20(b).

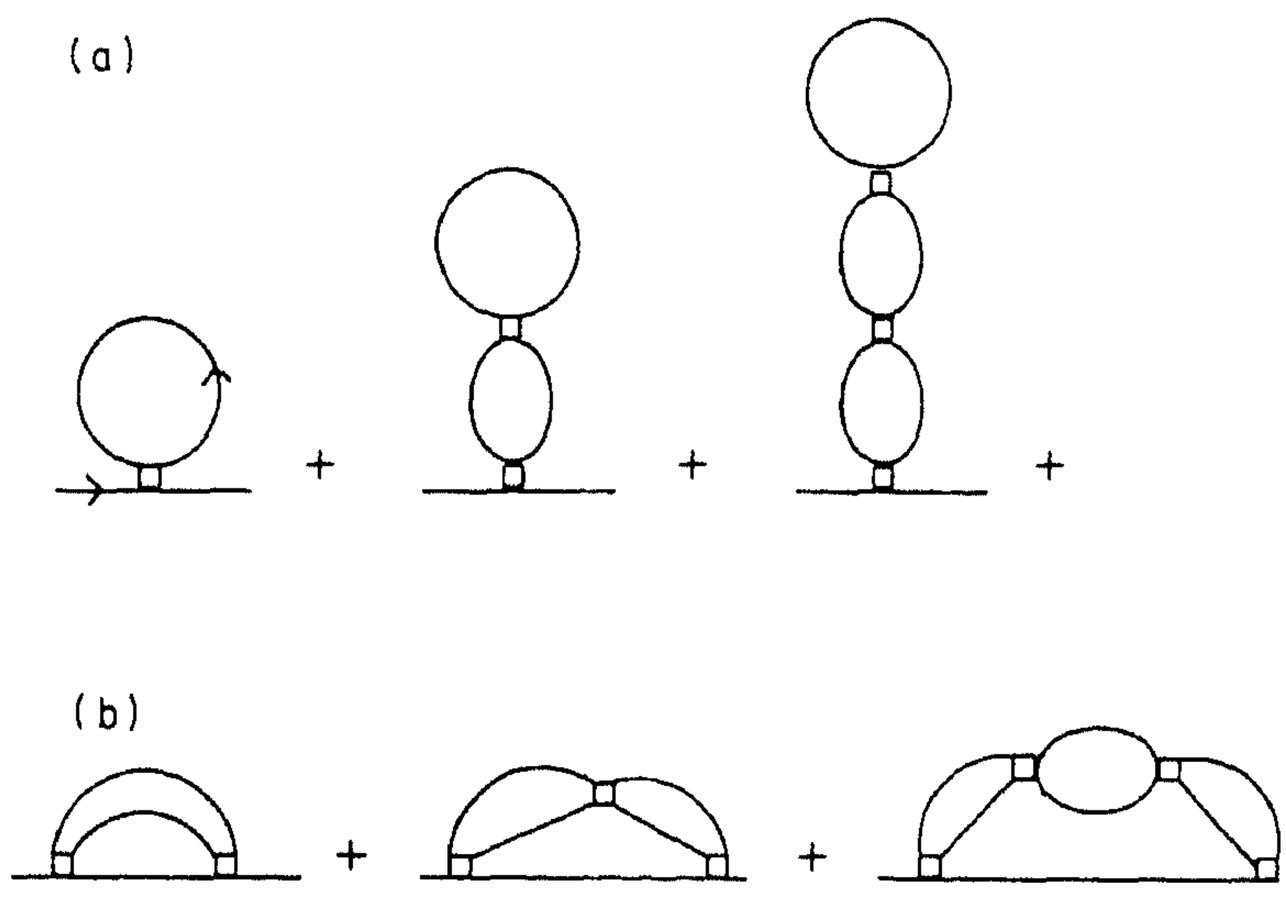


FIGURE 11.20

8. Use the paramagnon theory to study the effective interaction between quasiparticles. By summing the diagrams in Fig. 11.21, show that the effective interaction can be written as

$$V_{\text{eff}} = \frac{I}{2}\left(\frac{1}{1 - \frac{1}{2}IP^{(1)}} - \frac{\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'}{1 + \frac{1}{2}IP^{(1)}}\right) \tag{11.382}$$

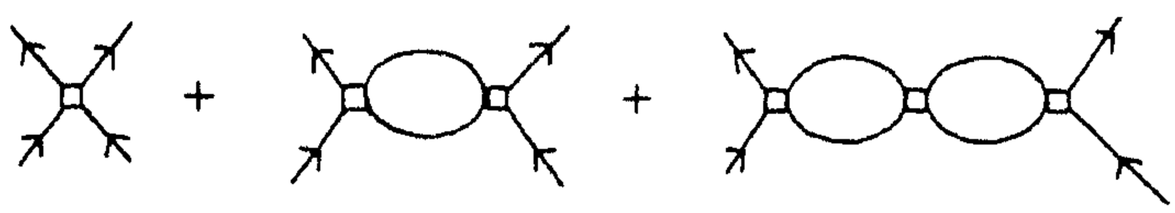


FIGURE 11.21

9. Calculate the quasiparticle self-energy in normal ^{3}He arising from the diagram in Fig. 11.22, and use the result to find the quasiparticle lifetime. Show that the result is (11.221) but without the factor $1 - n_1$.

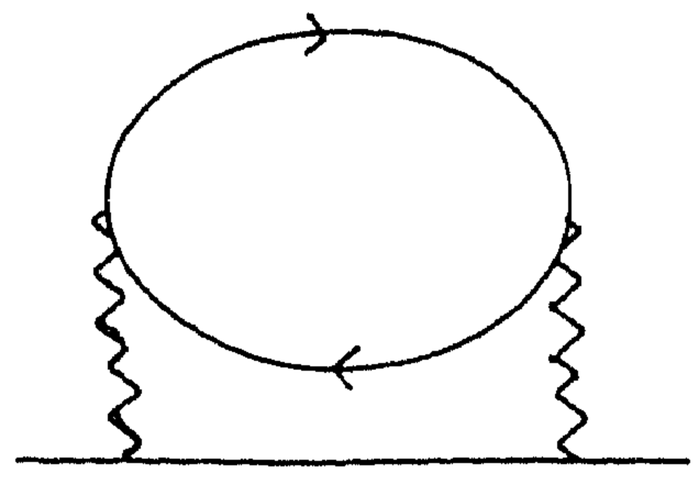


FIGURE 11.22

10. Use the result of Problem 7 [or Eq. (11.221)] to find the numerical value of the mean-free-path of a dressed quasiparticle in ^{3}He at $\xi_1 = 0$ at $T = 1$ mK.

11. Use Eqs. (11.209) to calculate the thermal conductivity and verify the result $k_B T = 33$ erg/cm-s in Table 11.4.

12. Find the zero-temperature gap function for the solution (11.266) to the gap equation for triplet pairing.

13. Show by explicit calculation that the choice (11.269) for the BW orbital alignment has the same gap function $\Delta(p)$ as the choice (11.256).

14. Calculate the pair distribution function $g(\mathbf{r}_1 - \mathbf{r}_2)$ in a system of only three particles using Eq. (11.331) for the case $\nu = 1$.

15. Use Eq. (11.331) for $N \gg 1$ to calculate the three-body distribution function $g(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ for the case that $\nu = 1$.

16. Use the Hartree–Fock wave function for the QHE with $\nu = 1$ to calculate the ODLRO $R(\rho, \rho')$.

17. In the QHE for $\nu = 1$, evaluate the partition function to derive the thermodynamic potential Ω_0.

18. Show that the Laughlin wavefunction can be written exactly as

$$\Psi = N\left[\Pi_{i>j} e^{-u(z_i - z_j)}\right]\Phi_{CM}\left(\sum_l z_l\right) \tag{11.383}$$

$$\Phi_{CM}(Z) = \exp\left[-\frac{1}{4N}|Z|^2\right] \tag{11.384}$$

$$u(z) = -p\ln(z) + \frac{1}{4N}|z|^2 \tag{11.385}$$

If there is no center of mass motion ($\sum_i z_i = 0$) then the wave function has exactly the Jastrow form. Use this form to show the density is uniform.

References

Chapter 1

ADLER, D., in *Solid State Physics*, Vol. 21, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1967), pp. 1–113.

ANDERSON, P. W., *Phys. Rev.* **124**, 41 (1961).

BORN, M., and K. HUANG, *Dynamical theory of Crystal Lattices* (Oxford University Press, London, 1954).

CALLEN, H., in *Physics of Many-Particle Systems*, Vol. 1, ed. E. Meeron (Gordon & Breach, New York, 1966), pp. 183–230.

DOMB, C., and M. S. GREEN, *Phase Transitions and Critical Phenomena*, Vols. 1–6 (Academic Press, New York, 1972–1977).

FANO, U., *Phys. Rev.* **124**, 1866 (1961).

FRöHLICH, H., H. PELZER, and S. ZIENAU, *Philos. Mag.* **41**, 221 (1950).

FRöHLICH, H., *Adv. Phys.* **3**, 325 (1954).

GERRITSEN, A. N., and J. O. LINDE, *Physica* **17**, 573, 584 (1951); **18**, 872 (1954).

GUTZWILLER, M. C., *Phys. Rev. Lett.* **10**, 159 (1963).

HARRISON, W. A., *Pseudopotentials in the Theory of Metals* (Benjamin, Reading, Mass., 1966).

HEINE, V., in *Solid State Physics*, Vol. 24, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1970).

HILL, T. L., *Statistical Mechanics* (McGraw-Hill, New York, 1956), Chap. 7.

HOPFIELD, J. J., *Phys. Rev.* **112**, 1555 (1958).

HUBBARD, J., *Proc. R. Soc. London Ser. A* **276**, 238 (1963); **277**, 237 (1964); **281**, 401 (1964); **285**, 542 (1965); **296**, 82, 100 (1966).

JORDAN, P., and E. WIGNER, *Z. Phys.* **47**, 631 (1928).

KONDO, J., *Prog. Theor. Phys. (Kyoto)* **32**, 37 (1964).

KONDO, J., in *Solid State Physics*, Vol. 23, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1969), pp. 183–281.

LIEB, E. H., and F. Y. WU, *Phys. Rev. Lett.* **20**, 1445 (1968).

MAHAN, G. D., *J. Phys. Chem. Solids* **26**, 751 (1965).

MAHAN, G. D., in *Polarons in Ionic Crystals and Polar Semiconductors*, ed. J. T. Devreese (North-Holland, Amsterdam, 1972), pp. 553–657.

MARADUDIN, A. A., E. W. MONTROLL, and G. H. WEISS, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, New York, 1963).

MATSUBARA, T., and H. MATSUDA, *Prog. Theor. Phys. (Kyoto)* **16**, 416 (1956).

MOZER, B., L. A. DEGRAAF, and B. LENEINDRE, *Phys. Rev. A* **9**, 448 (1974).

ONSAGER, L., *Phys. Rev.* **65**, 117 (1944).

PARR, R. G., *Quantum Theory of Molecular Electronic Structure* (Benjamin, Reading, Mass., 1964).

PERCUS, J. K., in *The Equilibrium Theory of Classical Fluids*, eds. H. L. Frisch and J. L. Lebowitz (Benjamin, Reading, Mass., 1964), pp. II-33–II-170.

SCHIFF, L. I., *Quantum Mechanics* (McGraw-Hill, New York, 1968).

SCHRIEFFER, J. R., and P. A. WOLFF, *Phys. Rev.* **149**, 491 (1966).
THOMAS, D. G., *J. Appl. Phys. Suppl.* **32**, 2298 (1961).
ZENER, C., *Phys. Rev.* **81**, 440 (1951).

Chapter 2

This reference list contains several books which have excellent descriptions of the material in this chapter.

ABRIKOSOV, A. A., L. P. GORKOV, and I. E. DZYALOSHINSKI, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963; Pergamon, Elmsford, N.Y., 1965).
AUERBACH, A., *Interacting Electronics and Quantum Magnetism* (Springer-Verlag, New York, 1994).
CRAIG, R., *J. Math. Phys* **9**, 605 (1968).
DONIACH, S., and E. H. SONDHEIMER, *Green's Functions for Solid State Physicists* (Benjamin, Reading, Mass., 1974).
FETTER, A. L., and J. D. WALECKA, *Quantum Theory of Many Particle Systems* (McGraw-Hill, New York, 1971).
GROSS, E. K. U., E. RUNGE, and O. HEINONEN, *Many-Particle Theory* (Hilger, New York, 1991).
GELL-MANN, M., and F. LOW, *Phys. Rev.* **84**, 350 (1951).
INKSON, J. C., *Many-Body Theory of Solids* (Plenum, New York, 1984).
KELDYSH, L., *Sov. Phys. JETP* **20**, 1018 (1965).
MATTOCK, R. D., *A Guide to Feynman Diagrams in the Many Body Problem* (McGraw-Hill, New York, 1967).
SCHWINGER, J., *J. Math. Phys.* **2**, 407 (1961).

Chapter 3

ABRIKOSOV, A. A., L. P. GORKOV, and I. E. DZYALOSHINSKI, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
BARNARD, R. D., *Thermoelectricity in Metals and Alloys* (Taylor & Francis, London, 1972).
BRANDT, W., and A. DUPASQUIER, editors of *Positron Solid State Physics,* Int. School "Enrico Fermi" (Varenna, Italy, 1981).
BROUT, R., and P. CARRUTHERS, *Lectures on the many-Electron Problem* (Wiley-Interscience, New York, 1963).
DE GROOT, S. R., *Thermodynamics of Irreversible Processes* (North-Holland, Amsterdam, 1952).
DE GROOT, S. R., and P. MAZUR, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
DUNN, D., *Can. J. Phys.* **53**, 321 (1975).
GREEN, M. S., *J. Chem. Phys.* **20**, 1281 (1952); **22**, 398 (1954).
KOHN, W., and J. M. LUTTINGER, *Phys. Rev.* **118**, 41 (1960).
KUBO, R., *Lectures in Theoretical Physics,* Vol. I (Boulder) (Wiley-Interscience, New York, 1959), pp. 120–203; *J. Phys. Soc. Japan* **12**, 570 (1957).
LANGRETH, D., in *NATO Advanced Study Institute on Linear and Nonlinear Transport,* ed. J. T. Devreese, V. E. van Doren (Plenum, New York, 1976), p. 3.
LEHMANN, H., *Nuovo Cimento* **11**, 342 (1954).
LUTTINGER, J. M., *Phys. Rev.* **135**, A1505 (1964).
LUTTINGER, J. M., and J. C. WARD, *Phys. Rev.* **118**, 1417 (1960).
MAHAN, G. D., *Phys. Rev. B* **11**, 4814 (1975).
MATSUBARA, T., *Prog. Theor. Phys. (Kyoto)* **14**, 351 (1955).
MILLS, A. P., W. S. CRANE, and K. F. CANTER, editors of *Positron Studies of Solids, Surfaces, Atoms* (World Scientific, Singapore, 1986).
TAYLOR, P. L., *A Quantum Approach to the Solid State* (Prentice-Hall, Englewood Cliffs, N.J., 1970).
TOMONAGA, S., *Prog. Theor. Phys. (Kyoto)* **2**, 6 (1947).
WIGNER, E., *Phys. Rev.* **40**, 749 (1932).

Chapter 4

AGRANOVICH, V. M., *Sov. Phys. JETP* **10**, 307 (1960); *Sov. Phys. Solid State* **3**, 592 (1961).
ALMBLADH, C. O., and P. MINNHAGEN, *Phys. Rev. B* **17**, 929 (1978).

ANDERSON, P. W., *Phys. Rev.* **124**, 41 (1961).
ANDERSON, P. W., and W. L. MCMILLAN, in *Theory of Magnetism in Transition Metals,* Course XXXVII, Enrico Fermi International School of Physics, Varenna, ed. W. Marshall (Academic Press, New York, 1967), pp. 50–86.
BAXTER, R. J., *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).
BORN, M., and K. HUANG, *Dynamical Theory of Crystal Lattices* (Oxford University Press, New York, 1954); see Fig. 18a on p. 91.
BRINKMAN, W. F., and T. M. RICE, *Phys. Rev. B* **2**, 1324 (1970).
BRUECKNER, K. A., C. A. LEVINSON, and H. M. MAHMOUD, *Phys. Rev.* **95**, 217 (1954).
CINI, M., and A. D'ANDREA, *J. Phys. C* **21**, 193 (1988).
COHEN, M. H., and F. KEFFER, *Phys. Rev.* **99**, 1128 (1955).
DAVYDOV, A. S., *Theory of Molecular Excitons* (Plenum, New York, 1971).
DEWITT, B. S., *Phys. Rev.* **103**, 1565 (1956).
DIETZ, R. E., D. G. THOMAS, and J. J. HOPFIELD, *Phys. Rev. Lett.* **8**, 391 (1962).
EDWARDS, S. F., *Philos. Mag.* **3**, 33, 1020 (1958).
ENGELSBERG, S., and B. B. VARGA, *Phys. Rev.* **136**, A1582 (1964).
FANO, U., *Phys. Rev.* **103**, 1202 (1956); **118**, 451 (1960).
FANO, U., *Phys. Rev.* **124**, 1866 (1961).
FEYNMAN, R. P., *Phys. Rev.* **84**, 108 (1951).
FRIEDEL, J., *Philos. Mag.* **43**, 153 (1952); *Adv. Phys.* **3**, 446 (1953).
FUKUDA, N., and R. G. NEWTON, *Phys. Rev.* **103**, 1558 (1956).
FUMI, F. G., *Philos. Mag.* **46**, 1007 (1955).
HEEGER, A., in *Chemistry and Physics of One-Dimensional Metals,* ed. H. J. Keller (Plenum, New York, 1977), pp. 87–135.
HENRY, C. H., and J. J. HOPFIELD, *Phys. Rev. Lett.* **15**, 964 (1965).
HOPFIELD, J. J., *Phys. Rev.* **112**, 1555 (1958).
KOHN, W., and J. M. LUTTINGER, *Phys. Rev.* **108**, 590 (1957).
LANGER, J. S., and V. AMBEGAOKAR, *Phys. Rev.* **121**, 1090 (1961).
LIEB, E. H., and D. C. MATTIS, *Mathematical Physics in One Dimension* (Academic Press, New York, 1966) Chap. 4.
LUTHER, A., and V. J. EMERY, *Phys. Rev. Lett.* **33**, 589 (1974).
LUTHER, A., and I. PESCHEL, *Phys. Rev. B* **9**, 2911 (1974).
LUTTINGER, J. M., *J. Math. Phys. N. Y.* **4**, 1154 (1963).
MAHAN, G. D., in *Electronics Structure of Polymers and Molecular Crystals,* eds. J. M. Andre and J. Ladik (Plenum Press, New York, 1975), pp. 79–158.
MAHAN, G. D., and F. H. CLARO, *Phys. Rev. B* **16**, 1168 (1977).
MATTIS, D. C., *The Many-Body Problem* (World Scientific, Singapore, 1993).
MATTIS, D. C., and E. H. LIEB, *J. Math. Phys.* **6**, 304 (1965).
OVERHAUSER, A. W., *Physics* **1**, 307 (1965).
SCHÖNHAMMER, K., and O. GUNNARSSON, *Solid State Commun.* **23**, 691 (1977).
SCHWEBER, S. S., *Relativistic Quantum Field Theory* (Harper & Row, New York, 1961).
TOMONAGA, S., *Prog. Theor. Phys. (Kyoto)* **5**, 544 (1950).
VAN HAERINGEN, W., *Phys. Rev.* **137**, A1902 (1965).
WILLIAMS, F. E., and M. H. HEBB, *Phys. Rev.* **84**, 1181 (1951).

Chapter 5

BARDEEN, J., *Phys. Rev.* **49**, 653 (1936).
BAYM, G., and L. P. KADANOFF, *Phys. Rev.* **124**, 287 (1961).
BETHE, H., *Ann. Phys.* **5**, 325 (1930).
BROSENS, F., L. F. LEMMENS, and J. T. DEVREESE, *Phys. Status Solidi* (b) **81**, 551 (1977).
CARE, C. M., and N. H. MARCH, *Adv. Phys.* **24**, 101 (1975).
CARR, W. J., *Phys. Rev.* **122**, 1437 (1961).
CARR, W. J., and A. A. MARADUDIN, *Phys. Rev.* **133**, A371 (1964).
DANIEL, E., and S. H. VOSKO, *Phys. Rev.* **120**, 2041 (1960).
EHRENREICH, H., and M. H. COHEN, *Phys. Rev.* **115**, 786 (1959).
EISENBERGER, P., L. LAM, P. M. PLATZMAN, and P. SCHMIDT, *Phys. Rev. B* **6**, 3671 (1972).

ECHENIQUE, P. M., J. M. PITARKE, E. V. CHULKOV, and A. RUBIO, *Chemical Physics* **251**, 1 (2000).
FERMI, E., *Z. Phys.* **48**, 73 (1928).
FLESZAR, A., A. A. QUONG, and A. G. EGUILUZ, *Phys. Rev. Lett.* **74**, 590 (1995).
FLODSTROM, S. A., R. Z. BACHRACH, R. S. BAUER, J. C. MCMENAMIN, and S. B. M. HAGSTROM, *J. Vac. Sci. Technol.* **14**, 303 (1977).
GELDART, D. J. W., and R. TAYLOR, *Can. J. Phys.* **48**, 155, 167 (1970).
GELL-MANN, M., and K. BRUECKNER, *Phys. Rev.* **106**, 364 (1957).
HAMMERBERG, J., and N. W. ASHCROFT, *Phys. Rev. B* **9**, 409 (1974).
HEDIN, L., *Phys. Rev.* **139**, A796 (1965).
HOPFIELD, J. J., *Phys. Rev. B* **2**, 973 (1970).
HUBBARD, J., *Proc. R. Soc. London Ser.* **A 243**, 336 (1957).
JONSON, M., *J. Phys. C* **9**, 3055 (1976).
KIMBALL, J. C., *Phys. Rev. A* **7**, 1648 (1973).
KRAMERS, H. A., reprinted in *Collected Scientific Papers* (North-Holland, Amsterdam, 1956).
KRAVCHENKO, S. V., G. V. KRAVCHENKO, J. E. FURNEAUX, V. M. PUDALOV, and M. D'IORIO, *Phys. Rev. B* **50**, 8039 (1994).
KRAVCHENKO, S. V., W. E. MASON, G. E. BOWKER, J. E. FURNEAUX, V. M. PUDALOV, and D'IORIO, *Phys. Rev. B* **51**, 7038 (1995).
KRONIG, R., *J. Opt. Soc. Am.* **12**, 547 (1926).
LANDAU, L. D., and E. LIFSHITZ, *Quantum Mechanics: Non Relativistic Theory* (Addison Wesley, Reading, Mass., 1958).
LANG, N. D., and W. KOHN, *Phys. Rev. B* **3**, 1215 (1971).
LANGER, J. S., and S. H. VOSKO, *J. Phys. Chem. Solids* **12**, 196 (1960).
LARSON, B. C., J. Z. TISCHLER, E. D. ISAACS, P. ZSCHACK, A. FLESZAR, and A. G. EQUILUZ, *Phys. Rev. Lett.* **77**, 1346 (1996).
LINDGREN, L., and A. ROSEN, unpublished (1970).
LINDHARD, J., *K. Dan. Vidensk. Selsk. Mat. Fys. Medd.* **28**, (8) (1954).
LUNDQVIST, B. L., *Phys. Status Solidi* **32**, 273 (1969).
MAHAN, G. D., and B. E. SERNELIUS, *Phys. Rev. Lett.* **62**, 2718 (1989).
MAHAN, G. D., *Mod. Phys. Lett.* **7**, 13 (1993).
MAO, H. K., and R. J. HEMLEY, *Rev. Mod. Phys.* **66**, 671 (1994).
MARTON, L., J. A. SIMPSON, H. A. FOWLER, and N. SWANSON, *Phys. Rev.* **126**, 182 (1962).
MIHARA, N., and R. D. PUFF, *Phys. Rev.* **174**, 221 (1968).
NIKLASSON, G., *Phys. Rev. B* **10**, 3052 (1974).
NOZIERES, P., *Interacting Fermi Systems* (Benjamin, Reading, Mass., 1964), p. 287.
NOZIERES, P., and D. PINES, *Phys. Rev.* **111**, 442 (1958).
ONSAGER, L., L. MITTAG, and M. J. STEPHEN, *Ann. Phys.* **18**, 71 (1966).
PANDEY, K. C., and L. LAM, *Phys. Lett.* **43A**, 319 (1973).
QUINN, J. J., and R. A. FERRELL, *Phys. Rev.* **112**, 812 (1958).
RICE, T. M., *Ann. Phys.* **31**, 100 (1965).
RITCHIE, R. H., *Phys. Rev.* **106**, 874 (1957).
SANTORO, G. E., and G. F. GIULIANI, *Phys. Rev. B* **37**, 4813 (1988).
SAWADA, K., K. A. BRUECKNER, N. FUKADA, and R. BROUT, *Phys. Rev.* **108**, 507 (1957).
SCHNEIDER, T., R. BROUT, H. THOMAS, and J. FEDER, *Phys. Rev. Lett.* **25**, 1423 (1970).
SEITZ, F., *Modern Theory of Solids* (McGraw-Hill, New York, 1940), Sec. 76.
SHAM, L. J., *Phys. Rev. B* **7**, 4357 (1973).
SHOLL, C. A., *Proc. Phys. Soc. London* **92**, 434 (1967).
SHUNG, K. W. K., B. E. SERNELIUS, and G. D. MAHAN, *Phys. Rev. B* **36**, 4499 (1987).
SINGWI, K. S., and M. P. TOSI, *Solid State Phys.* **36**, 177 (1981).
SINGWI, K. S., M. P. TOSI, R. H. LAND, and A. SJOLANDER, *Phys. Rev.* **176**, 589 (1968).
SINGWI, K. S., A. SJOLANDER, M. P. TOSI, and R. H. LAND, *Phys. Rev. B* **1**, 1044 (1970).
SUEOKA, O., *J. Phys. Soc. Jpn.* **20**, 2203 (1965).
THOMAS, L. H., *Proc. Cambridge Philos. Soc.* **23**, 542 (1927).
TING, C. S., T. K. LEE, and J. J. QUINN, *Phys. Rev. Lett.* **34**, 870 (1975).
TRACY, J. C., *J. Vac. Sci. Technol.* **11**, 280 (1974).
VASHISHTA, P., and K. S. SINGWI, *Phys. Rev. B* **6**, 875 (1972).
VIGNALE, G., *Phys. Rev. B* **38**, 6445 (1988).

WIGNER, E., *Phys. Rev.* **46**, 1002 (1934).
WIGNER, E., *Trans. Faraday Soc.* **34**, 678 (1938).
WIGNER, E., and E. SEITZ, *Phys. Rev.* **43**, 804 (1933); **46**, 509 (1934).

Chapter 6

ABRIKOSOV, A. A., *Physics* **2**, 61 (1965).
ANDERSON, P. W., *Phys. Rev.* **124**, 41 (1961).
ANDREI, N., K. FURUYA, and J. H. LOWENSTEIN, *Rev. Mod. Phys.* **55**, 331 (1983).
BARNES, S. E., *J. Phys. F* **6**, 1375 (1976); **7**, 2637 (1977).
BICKERS, N. E., *Rev. Mod. Phys.* **59**, 845 (1987).
BICKERS, N. E., D. L. COX, and J. W. WILKINS, *Phys. Rev. B* **36**, 2036 (1987).
CHEN, H., and D. C. MATTIS, *Int. J. Mod. Phys. B* **5**, 2951 (1991).
COLEMAN, P., *Phys. Rev. B* **29**, 3035 (1984).
GREWE, N., and H. KEITER, *Phys. Rev. B* **24**, 4420 (1981).
GUNNARSSON, O., and K. SCHÖNHAMMER, *Phys. Rev. B* **28**, 4315 (1983).
GUNNARSSON, O., and K. SCHÖNHAMMER, in *Handbook on the Physics and Chemistry of Rare Earths* ed. K. A. Gschneider, L. Eyring, and S. Hüfner (North-Holland, Amsterdam, 1987), Vol. 10, p. 103.
HERBST, J. F., and J. W. WILKINS, *Handbook of the Physics and Chemistry of Rare Earths*, Vol. 10, ed. K. A. Gschneider, L. Eyring, and S. Hüfner (Elsevier, New York, 1987), Chap. 68.
HEWSON, A. C., *The Kondo Problem to Heavy Fermions* (Cambridge University Press, Cambridge, 1993).
JACCARD, D., and J. SIERRO, in *Valence Instabilities*, ed. P. Wachter, and H. Boppart (North-Holland, Amsterdam, 1982), p. 409.
KANAMORI, J., *Prog. Theor. Phys.* **30**, 275 (1963).
KONDO, J., *Prog. Theor. Phys.* **32**, 37 (1964).
LEE, P. A., T. M. RICE, J. W. SERENE, L. J. SHAM, and J. W. WILKINS, *Comments on Condensed Matter Phys.* **XII**, 99–161 (1986).
LIU, S. H., *Phys. Rev. B* **37**, 3542 (1988).
MAHAN, G. D., in *Solid State Physics*, ed. H. Ehrenreich, and F. Spaen, Vol. 51 (Academic Press, New York, 1997), pp. 81–157.
MAHAN, G. D., and J. O. SOFO, *Proc. Natl. Acad. Sci. USA* **93**, 7436 (1996).
MERMIN. N. D., and H. WAGNER, *Phys. Rev. Lett.* **17**, 1133 (1966).
OVERHAUSER, A. W., *Phys. Rev. Lett.* **13**, 190 (1964).
OVERHAUSER, A. W., *Adv. Phys.* **27**, 343 (1978).
PENN, D. R., *Phys. Rev.* **142**, 350–365 (1966).
SAMWER, K., and K. WINZER, *Z. Phys. B* **25**, 269–274 (1976).
SCHRIEFFER, J. R., and P. A. WOLFF, *Phys. Rev.* **149**, 491 (1966).
SILVERSTEIN, S. D., and C. B. DUKE, *Phys. Rev.* **61**, 456, 470 (1967).
STEWART, G. R., *Rev. Mod. Phys.* **56**, 755 (1984).
VILK, Y. M., L. CHEN, and A. M. S. TREMBLAY, *Phys. Rev. B* **49**, 13267–13270 (1994).
VILK, Y. M., and A. M. S. TREMBLAY, *J. Phys. I* (France) **7**, 1309 (1997).
WIEGMANN, P., *J. Phys. C* **14**, 1463 (1981).
WILKINS, J. W., in *Valence Instabilities*, ed. P. Wachter, and H. Boppart (North-Holland, New York, 1982).
YOSIDA, K., *Phys. Rev.* **147**, 223 (1966).
ZWICKNAGL, G., *Adv. Phys.* **41**, 203 (1992).

Chapter 7

ADLER, D., in *Solid State Physics*. Vol. 21, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1967), p. 1.
ALLCOCK, G. R., *Adv. Phys.* **5**, 412 (1956).
ALLCOCK, G. R., in *Polarons and Excitons*, eds. C. G. Kuper, and G. D. Whitfield (Oliver & Boyd, Edinburgh, 1962), pp. 45–70.
APPEL, J., *Phys. Rev. Lett.* **17**, 1045 (1966).
APPEL. J., in *Solid State Physics*, Vol. 21, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1967), p. 193.

ASHCROFT, N. W., and J. W. WILKINS, *Phys. Lett.* **14**, 285 (1965).
BOGOMOLOV, V. N., E. K. KUDINOV, and Yu. A. FIRSOV, *Sov. Phys. Solid State* **9**, 2502 (1968).
BOHM, D., and I. STAVER, *Phys. Rev.* **84**, 836 (1951).
BOTTGER, H., and V. V. BRYKSIN, *Phys. Status Solidi B* **78**, 415 (1976).
BRILLOUIN, L., *J. Phys. Paris* **3**, 379 (1932); **4**, 1 (1933).
BROUT, R., and P. CARRUTHERS, *Lectures on the Many-Electron Problem* (Wiley-Interscience, New York, 1963).
BROWN, D. W., K. LINDENBERG, and Y. ZHAO, *J. Chem. Phys.* **107**, 3159, 3179 (1997).
CONLEY, J. W., and G. D. MAHAN, *Phys. Rev.* **161**, 681 (1967).
DEWIT, H. J., *Philips Res. Rep.* **23**, 449 (1968).
DUNN, D., *Can. J. Phys.* **53**, 321 (1975).
ENGELSBERG, S., and J. R. SCHRIEFFER, *Phys. Rev.* **131**, 993 (1963).
FEYNMAN, R. P., *Phys. Rev.* **97**, 660 (1955).
GARRETT, D. G., and J. C. SWIHART, *J. Phys. F* **6**, 1781 (1976).
GRIMVALL, G., *The Electron Phonon Interaction in Metals* (North-Holland, New York, 1981).
HARRISON, W. A., *Pseudopotentials in the Theory Of Metals* (Benjamin, Reading, Mass., 1966).
HODBY, J. W., in *Polarons in Ionic Crystals and Polar Semiconductors,* ed. J. Devreese (North-Holland, Amsterdam, 1972), pp. 389 –459.
HOLSTEIN, T., *Annals of Phys.* **8**, 343 (1959).
JOSHI, S. K., and A. K. RAJAGOPAL, in *Solid State Physics,* Vol. 22, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1968), pp. 160–313.
KARTHEUSER, E., R. EVRARD, and J. DEVREESE, *Phys. Rev. Lett.* **22**, 94 (1969).
KAY, P., and J. A. REISSLAND, *J. Phys. F* **6**, 1503 (1976).
KOHN, W., *Phys. Rev. Lett.* **2**, 395 (1959).
KUDINOV, R. K., D. N. MIRLIN, and Yu. A. FIRSOV, *Sov. Phys. Solid State* **11**, 2257 (1970).
LANDAU, L. D., and S. L. PEKAR, *Zh. Eksp. Teor. Fiz.* **16**, 341 (1946).
LANG, I. G., and Yu. A. FIRSOV, *Sov. Phys. JETP* **16**, 1301 (1963); *Sov. Phys. Solid State* **5**, 2049 (1964).
LOW, F., T. D. LEE, and D. PINES, *Phys. Rev.* **90**, 297 (1953).
LYDDANE, R. H., R. G. SACHS, and E. TELLER, *Phys. Rev.* **59**, 673 (1941).
MAHAN, G. D., *Phys. Rev.* **145**, 602 (1966).
MAHAN, G. D., in *Polarons in Ionic Crystals and Polar Semiconductors,* ed. J. Devreese (North-Holland, Amsterdam, 1972), pp. 553–657.
MCMILLAN, W. L., *Phys. Rev.* **167**, 331 (1968).
MIGDAL, A. B., *Sov. Phys. JETP* **7**, 996 (1958).
MIYAKE, S. J., *J. Phys. Soc. Jpn.* **41**, 745 (1976).
MOORADIAN, A., and G. B. WRIGHT, *Phys. Rev. Lett.* **16**, 999 (1966).
NAGAEV, E. L., *Sov. Phys. Solid State* **4**, 1611 (1963).
NAKAJIMA, S., and M. WATABE, *Prog. Theor. Phys.* **29**, 341 (1963).
REIK, H. G., in *Polarons in Ionic Crystals and Polar Semiconductors,* ed. J. Devreese (North-Holland, Amsterdam, 1972), pp. 679–714.
ROWELL, J. M., W. L. MCMILLAN, and W. L. FELDMANN, *Phys. Rev.* **178**, 897 (1969a); *Phys. Rev.* **180**, 658 (1969b); *Phys. Rev. B* **3**, 4065 (1971).
SCHIFF, L. I., *Quantum Mechanics*, 2nd ed. (McGraw-Hill, New York, 1955), p. 199.
SCHOTTE, K. D., *Z. Phys.* **196**, 393 (1966).
SCHRIEFFER, J. R., *Theory of Superconductivity* (Benjamin, Reading, Mass., 1964).
SHAM, L. J., and J. M. ZIMAN, *Solid State Phys.,* Vol. 15, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1963), p. 221.
SHUKLA, R. C., and R. TAYLOR, *J. Phys. F* **6**, 531 (1976).
SMONDYREV, M. A., *Theor. Math. Phys.* (Russian) **68**, 653 (1986).
SRIVASTAVA, P. L., and R. N. SINGH, *J. Phys. F* **6**, 1819 (1976).
SWIHART, J. C., J. SCALAPINO, and Y. WADA, *Phys. Rev. Lett.* **14**, 106 (1965).
TIABLIKOV, S. V., *Zh. Eksp. Teor. Fiz.* **23**, 381 (1952).
TOMLINSON, P. G., and J. P. CARBOTTE, *Solid State Commun.* **18**, 119 (1976).
TSUI, D. C., *Phys. Rev. B* **10**, 5088 (1974).
WIGNER, E. P., *Math. U. Naturw. Anz. Ungar. Akad. Wiss.* **53**, 477 (1935).
WOLL, E. J., and W. KOHN, *Phys. Rev.* **126**, 1693 (1962).
YAMASHITA, J., and T. KUROSAWA, *J. Phys. Chem. Solids* **5**, 34 (1958).
ZIMAN, J. M., *Electrons and Phonons* (Oxford University Press, Inc., New York, 1960).

Chapter 8

ALLEN, P. B., *Phys. Rev. B* **3**, 305 (1971).
ALLEN, P. B., *Phys. Rev. Lett.* **59**, 1460 (1987).
ARGYRES, P. N., and J. L. SIGEL, *Phys. Rev. B* **9**, 3197 (1974).
BASS, J., *Landolt-Bornstein*, Neue Serie, Band 15a (Springer-Verlag, New York, 1982).
CAROLI, C., R. COMBESCOT, D., LEDER-ROZENBLATT, P. NOZIERES, and D. SAINT JAMES, *Phys. Rev. B* **12**, 3977 (1975).
COHEN, M. H., L. M. FALICOV, and J. C. PHILLIPS, *Phys. Rev. Lett.* **8**, 316 (1962).
DYNES, R. C., and J. P. CARBOTTE, *Phys. Rev.* **175**, 913 (1968).
ENGELSBERG, S., and J. R. SCHRIEFFER, *Phys. Rev.* **131**, 993 (1963).
FEUCHTWANG, T. E., *Phys. Rev. B* **12**, 3979 (1975).
FISHMAN, R., *Phys. Rev. B* **39**, 2994 (1989).
FLEUROV, V. N., and A. N. KOZLOV, *J. Phys. F* **8**, 1899 (1978).
GIAEVER, I., and H. R. ZELLER, *Phys. Rev. Lett.* **20**, 1504 (1968).
GOLDHABER-GORDON, D., J. GORES, M. A. KASTNER, H. SHTRIKMAN, D. MAHALU, and U. MEIRAV, *Phys. Rev. Lett.* **1**, 5225 (1998).
GRIMVALL, G., *The Electron Phonon Interaction in Metals* (North-Holland, New York, 1981).
HODBY, J. W., in *Polarons in Ionic Crystals and Polar Semiconductors,* ed. J. Devreese (North-Holland, Amsterdam, 1972), pp. 389–459.
HOLSTEIN, T., *Ann. Phys. (N.Y.)* **29**, 410 (1964).
HOWARTH, D. J., and E. H. SONDHEIMER, *Proc. R. Soc. London Ser. A* **219**, 53 (1953).
HUBERMAN, M., and G. V. CHESTER, *Adv. Phys.* **24**, 489 (1975).
KADANOFF, L. P., and G. BAYM, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).
KITTEL, C., *Quantum Theory of Solids* (Wiley, New York, 1963) Chap. 9.
KUBO, R., M. TODA, and N. HASHITSUME, *Statistical Physics II* (Springer-Verlag, New York, 1985).
LANDAUER, R., *Phys. Lett.* **85A**, 91 (1981).
LANDAUER, R., *J. Phys.: Cond. Matter* **1**, 8099 (1989).
LANGER, J. S., *Phys. Rev.* **120**, 714 (1960).
LANGER, J. S., *Phys. Rev.* **124**, 997, 1003 (1961).
LANGRETH, D. C., *Phys. Rev.* **159**, 717 (1967).
LANGRETH, D. C., and L. P. KADANOFF, *Phys. Rev.* **133**, A1070 (1964).
LEVANDA, M., and V. N. FLEUROV, *J. Phys.: Cond. Matter* **6**, 789 (1994).
MAHAN, G. D., *Phys. Rev.* **142**, 366 (1966).
MAHAN, G. D., *J. Appl. Phys.* **58**, 2242 (1985).
MAHAN, G. D., *Phys. Rep.* **145**, 251 (1987).
MAHAN, G. D., and W. HANSCH, *J. Phys. F* **13**, L47 (1983); *Phys. Rev. B* **28**, 1902 (1983).
MATTHIESSEN, A., *Rep. Br. Assoc.* **32**, 144 (1862).
MIGDAL, A. B., *Sov. Phys. JETP* 7, 996 (1958).
OPSAL, J. L., B. J. THALER, and J. BASS, *Phys. Rev. Lett.* **36**, 1211 (1976).
PRANGE, R. E., and L. P. KADANOFF, *Phys. Rev.* **134**, A566 (1964).
SCHRIEFFER, J. R., D. J. SCALAPINO, and J. W. WILKINS, *Phys. Rev. Lett.* **10**, 336 (1963).
SEGALL, B., M. R. LORENZ, and R. E. HALSTAD, *Phys. Rev.* **129**, 2471 (1963).
TAKEGAHARA, K., and S. WANG, *J. Phys. F* **7**, L293 (1977).
TAYLOR, R., and A. H. MACDONALD, *Phys. Rev. Lett.* **57**, 1639 (1986).
WARD, J. C., *Phys. Rev.* **78**, 182 (1950).
ZAWADOWSKI, A., *Phys. Rev.* **163**, 341 (1967).
ZIMAN, J. M., *Electrons and Phonons* (Oxford University Press, Inc., New York, 1960).

Chapter 9

ALLEN, P. B., *Phys. Rev. B* **3**, 305 (1971).
ALMBLADH, C. O., *Phys. Rev. B* **16**, 4343 (1977).
ANDERSON, P. W., *Phys. Rev. Lett.* **18**, 1049 (1967).
ASHCROFT, N. W., and K. STURM, *Phys. Rev. B* **3**, 1898 (1971).

CALLCOTT, T. A., and E. T. ARAKAWA, *Phys. Rev. Lett.* **38**, 442 (1977).
CALLCOTT, T. A., E. T. ARAKAWA, and D. L. EDERER, *Phys. Rev. B* **18**, 6622 (1978).
CANRIGHT, G. S., *Phys. Rev. B* **38**, 1647 (1988).
CITRIN, P. H., *Phys. Rev. B* **8**, 5545 (1973).
CITRIN, P. H., G. K. WERTHEIM, and Y. BAER, *Phys. Rev. B* **16**, 4268 (1977).
DEVREESE, J. T., ed., in *Polarons in Ionic Crystals and Polar Semiconductors* (North-Holland, Amsterdam, 1972), pp. 92–93.
DONIACH, S., and M. SUNJIC, *J. Phys. C* **3**, 285 (1970).
ELLIOTT, R. J., *Phys. Rev.* **108**, 1384 (1957).
GALLON, T. E., in *Electron and Ion Spectroscopy of Solids,* eds. L. Fiermans, J. Vennik, and W. Dekeyser (Plenum Press, New York, 1978), pp. 230–272.
GIRVIN, S. M., and J. J. HOPFIELD, *Phys. Rev. Lett.* **37**, 1091 (1976).
GUPTA, R. P., and A. J. FREEMAN, *Phys. Rev. Lett.* **36**, 1194 (1976).
GUREVICH, V., I. LANG, and Yu. FIRSOV, *Sov. Phys. Solid State* **4**, 918 (1962).
HASEGAWA, M., and M. WATABE, *J. Phys. Soc. Jpn.* **27**, 1393 (1969).
HEDIN, L., and A. ROSENGREN, *J. Phys. F* **7**, 1339 (1977).
HOPFIELD, J. J., *Phys. Rev. A* **139**, 419 (1965).
HOPFIELD. J. J., unpublished (1967).
HOPFIELD, J. J., *Comments Solid State Phys.* **11**, 40 (1969).
KNOX, R. S., *Theory of Excitons* (Academic Press, New York, 1963), Supplement 5 to *Solid State Physics*, eds. H. Ehrenreich, F. Seitz, and D. Turnbull.
MAHAN, G. D., *Phys. Rev.* **153**, 882 (1967); **163**, 612 (1967).
MAHAN, G. D., *J. Phys. Chem. Solids* **31**, 1477 (1970).
MAHAN, G. D. *Solid State Physics*, Vol. 29, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1974), p. 75.
MAHAN, G. D., *Phys. Rev. B* **15**, 4587 (1977).
MAHAN, G. D., *Phys. Rev. B* **25**, 5021 (1982).
MINNHAGEN, P., *J. Phys. F* **7**, 2441 (1977).
MYERS, H. P., and P. SIXTENSSON, *J. Phys. F* **6**, 2023 (1976).
NOZIERES, P., and C. T. DEDOMINICIS, *Phys. Rev.* **178**, 1097 (1969).
OHMURA, Y., and A. OGIWARA, *Phil. Mag. B* **69**, 993–999 (1994).
OHTAKA, K., and Y. TANABE, *Phys. Rev. B* **28**, 6833 (1983); **34**, 3717 (1986).
PARDEE, W. J., and G. D. MAHAN, *Phys. Lett.* **45A**, 117 (1973).
PARDEE, W. J., G. D. MAHAN, D. E. EASTMAN, R. A. POLLAK, L. LEY, F. R. MCFEELY, S. P. KOWALCZYK, and D. A. SHIRLEY, *Phys. Rev. B* **11**, 3614 (1975).
PARRATT, L. G., *Rev. Mod. Phys.* **31**, 616, (1959).
PENN, D. R., S. M. GIRVIN, and G. D. MAHAN, *Phys. Rev. B* **24**, 6971 (1981).
REYNOLDS, D. C., and T. C. COLLINS, *Excitons: Their Properties and Uses* (Academic Press, New York, 1981).
SCHÖNHAMMER, K., and O. GUNNARSSON, *Phys. Rev. B* **18**, 6606 (1978).
SCHOTTE, K., and U. SCHOTTE, *Phys. Rev.* **182**, 479 (1969).
SIMÁNMEK, E., *Phys. Lett. A* **37**, 175 (1971).
SMITH, N. V., *Phys. Rev.* **183**, 634 (1969); *Phys. Rev. B* **2**, 2840 (1970).
WANNIER, G. H., *Phys. Rev.* **52**, 191 (1937).
WERTHEIM, G. K., D. M. RIFFE, and P. H. CITRIN, *Phys. Rev. B* **45**, 8703 (1992).
WU, J. W., and G. D. MAHAN, *Phys. Rev. B* **29**, 1769 (1984).

Chapter 10

ABRIKOSOV, A. A., L. P. GORKOV, and I. E. DZALOSHINSKI, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
ALLEN, P. B., M. L. COHEN, and D. R. PENN, *Phys. Rev. B* **38**, 2513 (1988).
ALLEN, P. B., and B. MITROVIĆ, in *Solid State Physics*, Vol. 37, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1982), pp. 1–82.
AMBEGAOKAR, V., and A. BARATOFF, *Phys. Rev. Lett.* **10**, 486 (1962); (E) **11**, 104 (1963).
BAKER, G. A., Jr., *Essentials of Padé Approximants* (Academic Press, New York, 1975).
BALIAN, R., and N. R. WERTHAMER, *Phys. Rev.* **131**, 1553 (1963).

BARDEEN, J., L. N. COOPER, and J. R. SCHRIEFFER, *Phys. Rev.* **108**, 1175 (1957).
COOPER, L. N., *Phys. Rev.* **104**, 1189 (1956).
ELIASHBERG, G. M., *Sov. Phys. JETP* **11**, 696 (1960).
FISKE, M. D., unpublished (1978).
FRÖHLICH, H., *Phys. Rev.* **79**, 845 (1950).
GIAEVER, I., *Phys. Rev. Lett.* **5**, 147, 464 (1960).
GIAEVER, I., H. R. HART, and K. MEGERLE, *Phys. Rev.* **126**, 941 (1962).
HALPERIN, B. I., and T. M. RICE, in *Solid State Physics*, Vol. 21, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1968), p. 115.
Han, S., K. W. NG, E. L. WOLF, A. MILLIS, J. L. SMITH, and Z. FISK, *Phys. Rev. Lett.* **57**, 238 (1986).
JOSEPHSON, B. D., *Phys. Lett.* **1**, 251 (1962).
KELDYSH, L. V., and Y. V. KOPAEV, *Sov. Phys. Solid State* **6**, 2219 (1965).
KITTEL, C., *Introduction to Solid State Physics*, 3rd ed. (Wiley, New York, 1966), p. 458.
LANGENBURG, D. N., D. J. SCALAPINO, B. N. TAYLOR, and R. E. ECK, *Phys. Rev. Lett.* **15**, 294 (1965).
MAHAN, G. D., *Phys. Rev.* **153**, 882 (1967); **163**, 612 (1967).
MAHAN, G. D., *Phys. Rev. B* **56**, 8322–8329 (1997).
MATTIS, D. C., and J. BARDEEN, *Phys. Rev.* **111**, 412 (1958).
MAXWELL, E., *Phys. Rev.* **78**, 477 (1950).
MCMILLAN, W. L., *Phys. Rev.* **167**, 331 (1968).
MCMILLAN, W. L., and J. M. ROWELL, *Phys. Rev. Lett.* **14**, 108 (1965).
MIGDAL, A. B. *Sov. Phys. JETP* **7**, 996 (1958).
MORSE, R. W., and H. V. BOHM, *Phys. Rev.* **108**, 1094 (1957).
OWEN, C. S., and D. J. SCALAPINO, *Physica* **55**, 691 (1971).
PALMER, L. H., and M. TINKHAM, *Phys. Rev.* **165**, 588 (1968).
PARKER, W. H., D. N. LANGENBERG, A. DENENSTEIN, and B. N. TAYLOR, *Phys. Rev.* **177**, 639 (1969).
PARKS, R. D., ed., *Superconductivity,* Vols. I & II (Dekker, New York, 1969).
REYNOLDS, C. A., B. SERIN, W. H. WRIGHT, and L. B. NESBITT, *Phys. Rev.* **78**, 487 (1950).
RICKAYZEN, G., *Theory of Superconductivity* (Wiley-Interscience, New York, 1965).
SCALAPINO, D. J., J. R. SCHRIEFFER, and J. W. WILKINS, *Phys. Rev.* **148**, 263 (1966).
SCHAFROTH, M. R., *Phys. Rev.* **100**, 463 (1955).
SCHILLING, A., M. CANTONI, J. D. GUO, and H. R. OTT, *Nature,* **363**, 56 (1993).
SCHRIEFFER, J. R., *Theory of Superconductivity* (Benjamin, Reading, Mass., 1964).
SHAW, W., and J. C. SWIHART, *Phys. Rev. Lett.* **20**, 1000 (1968).
SHEKHTMAN, L., and L. I. GLAZMAN, *Phys. Rev. B* **52**, R2297 (1995).
VAN DER LAGE, F. C., and H. A. BETHE, *Phys. Rev.* **71**, 612 (1947).
VIDBERG, H. J., and J. W. SERENE, *J. Low Temp. Phys.* **29**, 179 (1977).
WOLLMAN, D. A., D. J. VAN HARLINGEN, J. GIAPINTZAKIS, and D. M. GINSBERG, *Phys. Rev. Let.* **74**, 797 (1995).

Chapter 11

ABRIKOSOV, A. A., and I. M. KHALATNIKOV, *Rep. Prog. Phys.* **22**, 329 (1959).
ACHTER, E. K., and L. MEYER, *Phys. Rev.* **188**, 291 (1969).
ANDERSON, P. W., and W. F. BRINKMAN, in *The Helium Liquids*, eds. J. G. M. Armitage, and I. E. Farquhar (Academic Press, New York, 1975), pp. 315–416.
BALIAN, R., and N. R. WERTHAMER, *Phys. Rev.* **131**, 1553 (1963).
BAYM, G., and C. PETHICK, in *The Physics of Liquid and Solid Helium*, Part 2, eds. J. B. Ketterson, and K. Bennemann (Wiley, New York, 1978), Chap. 1.
BERK, N. F., and J. R. SCHRIEFFER, *Phys. Rev. Lett.* **17**, 433 (1966).
BIJL, A., *Physica* **7**, 860 (1940).
BOGOLIUBOV, N. N., *J. Phys. Moscow* **11**, 23 (1947).
BOGOLIUBOV, N. N., and D. N. ZUBAREV, *Sov. Phys. JETP* **1**, 83 (1955).
BROOKER, G. A., and J. SYKES, *Phys. Rev. Lett.* **21**, 279 (1968).
BROWN, G. V., and M. H. COOPERSMITH, *Phys. Rev.* **178**, 327 (1969).
COWLEY, R. A., and A. D. B. WOODS, *Can. J. Phys.* **49**, 177 (1971).
DINGLE, R. M., *Philos. Mag.* **40**, 573 (1949).

DONIACH, S., and S. ENGELSBERG, *Phys. Rev. Lett.* **17**, 750 (1966).
DY, K. S., and C. J. PETHICK, *Phys. Rev.* **185**, 373 (1969).
FEENBERG, E., *Theory of Quantum Fluids* (Academic Press, New York, 1969).
FEYNMAN, R. P., *Phys. Rev.* **94**, 262 (1954).
FEYNMAN, R. P., and M. COHEN, *Phys. Rev.* **102**, 1189 (1956); **107**, 13 (1957).
FRANCIS, W. P., G. V. CHESTER, and L. REATTO, *Phys. Rev. A* **1**, 86 (1970).
GIRVIN, S. M., A. H. MACDONALD, and P. M. PLATZMAN, *Phys. Rev. B* **33**, 2481 (1986).
GIRVIN, S. M., in *The Quantum Hall Effect,* eds. R. E. Prange, and S. M. Girvin (Springer-Verlag, New York, 1987).
GLYDE, H. R., *Excitations in Liquid and Solid Helium* (Oxford University Press, Oxford, 1994).
GREYWALL, D. S., *Phys. Rev. B* **27**, 2747 (1983).
GREYWALL, D. S., *Phys. Rev. B* **33**, 7520 (1986).
GRIFFIN, A., *Phys. Rev. B* **32**, 3289 (1985).
HALLOCK, R. B., *J. Low Temp. Phys.* **9**, 109 (1972).
HARLING, O., *Phys. Rev. A* **3**, 1073 (1971).
HIRSCHFELDER, J. O., C. F. CURTISS, and R. B. BIRD, *Molecular Theory of Gasses and Liquids* (Wiley, New York, 1954), p. 110.
HOHENBERG, P. C., and P. M. PLATZMAN, *Phys. Rev.* **152**, 198 (1966).
HONE, D., *Phys. Rev.* **121**, 669 (1961).
JAIN, J. K., *J. Phys. Chem. Solids* **51**, 889 (1990).
JANCOVICI, B., *Phys. Rev. Lett.* **46**, 386 (1981).
JASTROW, R., *Phys. Rev.* **98**, 1479 (1955).
JENSEN, H. H., H. SMITH, and J. W. WILKINS, *Phys. Lett. A* **27**, 532 (1968).
KLITZING, Von K., G. DORDA, and M. PEPPER, *Phys. Rev. Lett.* **45**, 494 (1980).
KOHN, W., *Phys. Rev.* **123**, 1242 (1961).
KOHN, W., and D. SHERRINGTON, *Rev. Mod. Phys.* **42**, 1 (1970).
KOSTERLITZ, J. M., and D. J. THOULESS, *J. Phys. C* **6**, 1181 (1973).
LAUGHLIN, R. B., *Phys. Rev. Lett.* **50**, 1395 (1983).
LANDAU, L. D., *J. Phys. Moscow* **5**, 71 (1941); **11**, 91 (1947); translations in I. M. Khalatnikov, *Introduction to the Theory of Superfluidity* (Benjamin, Reading, Mass., 1965).
LANDAU, L. D., *J. Phys. Moscow* **10**, 25 (1946); *Sov. Phys. JETP* **3**, 920 (1956); **5**, 101 (1957).
LEE, D. M., D. D. OSHEROFF, and R. C. RICHARDSON, *Rev. Mod. Phys.* **69**, 645–690 (1997).
LEGGETT, A. J., *Phys. Rev.* **140**, A1869 (1965).
LEGGETT, A. J., *Rev. Mod. Phys.* **47**, 331 (1975).
LEVIN, K., and O. T. VALLS, *Phys. Rev. B* **17**, 191 (1978).
LIFSHITS, E. M., and E. L. ANDRONIKASHIVILE, *A Supplement to Helium* (Consultants Bureau, New York, 1959), p. 75.
LILLY, M. P., K. B. COOPER, J. P. EISENSTEIN, L. N. PFEIFFER, and K. W. WEST, *Phys. Rev. Lett.* **82**, 394 (1999).
LUTTINGER, J. M., and P. NOZIERES, *Phys. Rev.* **127**, 1423, 4131 (1962).
MASSEY, W. E., *Phys. Rev.* **151**, 153 (1966).
MASSEY, W. E., and C. W. WOO, *Phys. Rev.* **164**, 256 (1967).
MCMILLAN, W. L., *Phys. Rev.* **138**, A442 (1965).
MOOK, H. A., *Phys. Rev. Lett.* **55**, 2452 (1985).
OSHEROFF, D. D., R. C. RICHARDSON, and D. M. LEE, *Phys. Rev. Lett.* **28**, 885 (1972).
OSHEROFF, D. D., W. J. GULLY, R. C. RICHARDSON, and D. M. LEE, *Phys. Rev. Lett.* **29**, 920 (1972).
PATTON, B. R., and A. ZARINGHALAM, *Phys. Lett. A* **55**, 95 (1975).
PENROSE, O., *Philos. Mag.* **42**, 1373 (1951).
PENROSE, O., and L. ONSAGER, *Phys. Rev.* **104**, 576 (1956).
PERCUS, J. K., in *The Equilibrium Theory of Classical Fluids*, eds. H. L. Frisch and J. L. Lebowitz (Benjamin, Reading, Mass., 1964), pp. 11-33–11-170.
PINES, D., and P. NOZIERES, *The Theory of Quantum Liquids* (Benjamin, Reading, Mass., 1966).
PRANGE, R. E., in *The Quantum Hall Effect*, eds. R. E. Prange and S. M. Girvin (Springer-Verlag, New York, 1987).
PUFF, R. D., and J. S. TENN, *Phys. Rev. A* **1**, 125 (1970).
REATTO, L., and G. V. CHESTER, *Phys. Rev.* **155**, 88 (1967).
RICE, M. J., *Phys. Rev.* **159**, 153 (1967); **162**, 189 (1967); **163**, 206 (1967).
RODRIQUEZ, L. J., H. A. GERSCH, and H. A. MOOK, *Phys. Rev. A* **9**, 2085 (1974).
RON, A., in *The Helium Liquids*, eds. J. G. M. Armitage and L. E. Farquhar (Academic Press, New York, 1975), pp. 211–240.

RUVALDS, J., and A. ZAWADOWSKI, *Phys. Rev. Lett.* **25**, 333, 632 (1970).
SKöLD, K., and C. A. PELIZZARI, *Philos. Trans. Roy. Soc. (London) B* **290**, 605 (1980).
STIRLING, W. G., and H. R. GLYDE, *Phys. Rev. B* **41**, 4224 (1990).
STÖRMER, H. L., (unpublished) (1999).
SVENSSON, E. C., V. F. SEARS, and A. GRIFFIN, *Phys. Rev. B* **23**, 4493 (1981).
SVENSSON, E. C., and V. F. SEARS, *Physica* **137B**, 126 (1986).
TSUI, D. C., H. L. STöRMER, and A. C. GOSSARD, *Phys. Rev. Lett.* **48**, 1559 (1982).
VOLLHARDT, D., and P. WöLFE, *The Superfluid Phases of Helium 3* (Taylor & Francis, London, 1990).
WHEATLEY, J. C., *Rev. Mod. Phys.* **47**, 415 (1975).
WOODS, A. D. B., and V. F. SEARS, *Phys. Rev. Lett.* **39**, 415 (1977).
YANG, C. N., *Rev. Mod. Phys.* **34**, 694 (1962).
YENNIE, D. R., *Rev. Mod. Phys.* **59**, 781 (1987).
ZAWADOWSKI, A., J. RUVALDS, and J. SOLNA, *Phys. Rev. A* **5**, 399 (1972).

Author Index

Abrikosov, A. A., 142, 148, 391, 635, 640, 729, 730, 766, 769, 772, 773
Achter, E. K., 722, 773
Adler, D., 23, 463, 765, 769
Agranovich, V. M., 285, 766
Allcock, G. R., 448, 769
Allen, P. B., 531, 542, 591, 592, 628, 671, 672, 674, 771, 772
Almbladh, C. O., 247, 606, 766, 771
Ambegaokar, V., 197, 207, 652, 655, 767, 772
Anderson, P. W., 44, 208, 214, 389, 603, 612, 675, 742, 765, 767, 769, 771, 773
Andrei, N., 387, 769
Andronikashivile, E. L., 710, 774,
Appel, J., 463, 497, 769, 770
Arakawa, E. T., 606, 772
Argyres, P. N., 536, 771
Ashcroft, N. W., 316, 491, 590, 768, 770, 772
Auerbach, A., 766

Bachrach, R. Z., 768
Baer, Y., 772
Baker, G. A., 668, 772
Balian, R., 635, 735, 736, 740, 772, 773
Baratoff, A., 652, 655, 772
Bardeen, J., 302, 627, 660, 662, 767, 772, 773
Barnard, R. D., 178, 181, 766
Barnes, S. E., 391, 769
Bass, J., 535, 771
Bauer, R. S., 768
Baxter, R., 187, 767
Baym, G., 337, 550, 554, 714, 767, 771, 773
Berk, N. F., 728, 773
Best, F. E., 333
Bethe, H., 369, 666, 767, 773
Bickers, N. E., 391, 396, 401, 769
Bijl, A., 690, 702, 703, 742, 773
Bird, R. B., 774
Bogoliubov, N. N., 678, 682, 694, 773
Bogomolov, V. N., 463, 770
Bohm, D., 486, 642, 770, 773
Bottger, H., 463, 770
Born, M., 11, 276, 765, 767
Bowker, G. E., 768
Billouin, L., 434, 770
Brandt, W., 168, 766
Brinkman, W. F., 247, 742, 767, 773
Brooker, G. A., 729, 773
Brosens, F., 337, 767
Brout, R., 142, 152, 448, 766, 768, 770
Brown, D. W., 456, 770
Brown, G. V., 686, 773
Brueckner, K. A., 204, 297, 300, 307, 363, 767, 768
Bryksin, V. V., 463, 770

Callcott, T. A., 606, 622, 623, 772
Callen, H., 38, 39, 765
Canter, K. F., 766
Canright, G. S., 607, 772
Cantoni, M., 773
Carbotte, J. P., 494, 524, 537, 770, 771
Care, C. M., 313, 767
Caroli, C., 561, 771
Carr, W. J., 307, 313, 767
Carruthers, P., 142, 152, 448, 766, 770
Chen, H., 405, 769
Chen, L., 769
Chester, G. V., 536, 694, 771, 774
Chulkov, E. V., 768
Cini, M., 247, 767
Citrin, P. H., 605, 625, 769, 772
Claro, F. H., 254, 767
Cohen, M., 709, 774
Cohen, M. H., 286, 325, 561, 767, 771
Cohen, M. L., 772
Coleman, P., 391, 769
Collins, T. C., 596, 772
Combescot, R., 771
Conley, J. W., 480, 770
Cooper, K. B., 774
Cooper, L. N., 627, 628, 629, 773
Coppersmith, M. H., 686, 773
Cowley, R. A., 702, 773
Cox, D. L., 769
Craig, R., 107, 766
Crane, W. S., 766
Curtiss, C. F., 774

D'Andrea, A., 247, 767
Daniel, E., 365, 366, 767
Davydov, A. S., 282, 767
Dedominicis, C. T., 603, 611, 613, 621, 772
De Groot, S. R., 177, 178, 766
Degraaf, L. A., 765
Denenstein, A., 773
Devreese, J. T., 584, 585, 767, 770 772
Dewit, H. J., 461, 770
Dewitt, B. S., 204, 767
Dietz, R. E., 236, 237, 767
Dingle, R. M., 690, 774
D'Iorio, M., 768
Domb, C., 38, 765
Doniach, S., 603, 624, 626, 729, 766, 772, 774
Dorda, G., 774
Duke, C. B., 387, 769
Dunn, D., 152, 44, 451, 452, 453, 766, 770
Dupasquier, A.. 168, 766

Dy, K. S., 732, 733, 734, 774
Dynes, R. C., 524, 537, 771
Dzyaloshinski, I. E., 766, 772

Eastman, D. E., 772
Echenique, P. M., 370, 768
Eck, R. E., 773
Ederer, D. L., 772
Edwards, S. F., 200, 767
Eguliuz, A. G., 768
Ehrenreich, H., 325, 767
Eisenberger, P., 366, 367, 368, 767
Eisenstein, J. P., 774
Eliashberg, G. M., 627, 664, 667, 773
Elliott, R. J., 594, 595, 600, 772
Emery, V. J., 256, 767
Engelsberg, S., 246, 480, 515, 516, 539, 729, 767, 770, 771, 774
Evrard, R., 770

Falicov, L. M., 771
Fano, U., 44, 208, 276, 765, 767
Feder, J., 768
Feenberg, E., 707, 708, 709, 713, 721, 774
Feldmann, W. L., 770
Fermi, E., 323, 768
Fetter, A. L., 766
Ferrell, R. A., 300, 301, 302, 363, 768
Feynman, R. P., 81, 220, 434, 438, 446, 697, 702, 703, 709, 742, 767, 770, 774
Feuchtwang, T. E., 561, 771
Firsov, Yu., A., 463, 770, 772
Fishman, R., 536, 771
Fisk, Z., 773
Fiske, M. D., 658, 773
Fleurov, V. N., 553, 554, 771
Fleszar, A., 342, 768
Flodstrom, S. A., 370, 768
Fowler, H. A., 768
Francis, W. P., 690, 691, 694, 695, 696, 774
Freeman, A. J., 623, 772
Friedel, J., 196, 204, 767
Frohlich, H., 26, 34, 627, 765, 773
Fukada, N., 204, 767, 768
Fumi, F. G., 198, 204, 767
Furneaux, J. E., 768
Furuya, K., 769

Gallon, T. E., 605, 772
Garrett, D. G., 487, 770
Geldart, D. J. W., 330, 768
Gell-Mann, M., 71, 297, 300, 307, 363, 766, 768
Gerritsen, A. N., 46, 765
Gersch, H. A., 775
Giaever, I., 567, 628, 646, 647, 650, 651, 660, 771, 773
Giapintzakis, J., 773
Ginsberg, D. M., 773
Girvin, S. M., 623, 626, 757, 758, 759, 761, 772, 774
Giuliani, G. F., 341, 358, 768
Glazman, L. I., 652, 773
Glyde, H. R., 703, 715, 723, 774, 775
Goldhaber-Gordon, D., 575, 771
Gores, J., 771
Gorkov, L. P., 766, 772
Gossard, A. C., 775
Green, M. S., 38, 160, 765, 766
Grewe, N., 391, 769
Greywall, D. S., 720, 735, 774
Griffin, A., 694, 774, 775
Grimvall, G., 491, 492, 545, 770, 771
Gross, E. K. U., 766
Gully, W. J., 774
Gunnarsson, O., 214, 391, 398, 614, 767, 769, 772
Guo, J. D., 773
Gupta, R. P., 623, 772
Gurevich, V., 585, 587, 772
Gutzwiller M. C., 22, 765

Hagstrom, S. B. M., 768
Hallock, R. B., 722, 774
Halperin, B. I., 634, 773
Halstad, R. E., 771
Hammerberg, J., 316, 768
Han, S., 635, 773
Hansch, W., 535, 550, 771
Harling, O., 698, 703, 774
Harrison, W. A., 19, 482, 487, 765, 770
Hart, H. R., 773
Hasegawa, M., 582, 772
Hashitsume, N., 771
Hebb, M. H., 232, 767
Hedin, L., 366, 606, 768, 772
Heeger, A., 256, 767
Heine, V., 19, 765
Heinonen, O., 766
Hemley, R. J., 315, 768
Henry, C. H., 278, 279, 767
Herbst, J. F., 391, 392, 393, 769
Hewson, A. C., 391, 769
Hill, T. L., 43, 58, 765
Hirschfelder, J. O., 678, 774
Hodby, J. W., 447, 517, 770, 771
Hohenberg, P. C., 697, 774
Holstein, T., 62, 401, 455, 457, 463, 499, 535, 537, 543, 770, 771
Hone, D., 729, 730, 774
Hopfield, J. J., 52, 276, 278, 279, 361, 584, 612, 620, 623, 626, 765, 767, 768, 772
Howarth, D. J., 519, 771
Huang, K., 11, 276, 765, 767
Huberman, M., 536, 771
Hubbard, J., 22, 300, 336, 338, 765, 768

Inkson, J. C., 766
Isaacs, E. D., 768

Jaccard, D., 402, 769
Jain, J. K., 761, 774
Jancovici, B., 754, 774
Jastrow, R., 690, 774
Jensen, H. H., 729, 774
Jonson, M., 374, 768
Jordan, P., 17, 40, 765
Josephson, B. D., 644, 656, 773
Joshi, S. K., 486, 770

Kadanoff, L. P., 337, 519, 545, 550, 554, 767, 771
Kanamori, J., 413, 427, 769
Kartheuser, E., 448, 770
Kastner, M. A., 771
Kay, P., 487, 770
Keffer, F., 286, 767
Keiter, H., 391, 769
Keldysh, L., 107, 634, 766, 773
Khalatnikov, I. M., 729, 730, 773
Kimball, J. C., 346, 768
Kittel, C., 500, 642, 771, 773
Klitzing, Von K., 742, 774
Knox, R. S., 596, 772
Kohn, W., 150, 199, 302, 487, 687, 758, 766, 767, 768, 770, 774
Kondo, J., 43, 46, 375, 765, 769
Kopaev, Y. V., 634, 773
Kosterlitz, J. M., 761, 774
Kowalczyk, S. P., 772
Kozlov, A. N., 554, 771
Kramers, H. A., 360, 768
Kravchenko, G. V., 768
Kravchenko, S. V., 356, 768
Kronig, R., 360, 768
Kubo, R., 160, 179, 535, 536, 766, 771
Kudinov, E. K., 465, 770
Kunz, C., 333
Kurosawa, T., 455, 770

Lam, L., 367, 767, 768
Land, R. H., 768
Landau, L. D., 324, 444, 703, 710, 711, 713, 768, 770, 774

Landauer, R., 575, 771
Lang, I. G., 463, 770, 772
Lang, N. D., 302, 768
Langenburg, D. N., 656, 773
Langer, J. S., 197, 207, 347, 505, 515, 516, 767, 768
Langmuir, I., 332
Langreth, D., 156, 519, 766, 771
Larson, B. C., 342, 345, 768
Laughlin, R. B., 752, 760, 774
Leder-Rozenblatt, D., 771
Lee, D. M., 735, 774
Lee, P. A., 390, 769
Lee, T. D., 770
Lee, T. K., 768
Leggett, A. J., 714, 727, 741, 742, 774
Lehmann, H., 122, 766
Lemmens, L. F., 767
LeNeindre, B., 765
Levanda, M., 553, 771
Levin, K., 729, 774
Levinson, C. A., 767
Ley, L., 772
Lieb, E. H., 23, 187, 267, 282, 765, 767
Lifshitz, E., 324, 710, 768, 774
Lilly, M. P., 761, 774
Linde, J. O., 765
Lindenberg, K., 770
Lindgren, L., 313, 768
Lindhard, J., 325, 768
Liu, S. H., 401, 403, 769
Lorenz, M. R., 771
Low, F., 71, 434, 766, 770
Lowenstein, J. H., 769
Lundqvist, B. L., 306, 313, 369, 370, 768
Luther, A., 256, 267, 269, 767
Luttinger, J. M., 150, 170, 178, 199, 715, 766, 767, 774
Lyddane, R. H., 472, 475, 770

MacDonald, A. H., 548, 549, 771
Mahalu, D., 771
Mahan, G. D., 32, 34, 152, 254, 282, 285, 344, 364, 401, 402, 436, 448, 450, 451, 471, 480, 520, 523, 534, 535, 550, 554, 582, 592, 603, 606, 607, 610, 620, 621, 634, 675, 765, 766, 768, 769, 770, 773
Mahmoud, H. M., 767
Maradudin, A. A., 11, 307, 313, 765, 767
March, N. H., 313, 767
Marton, L., 370, 371, 372, 768
Mason. W. E., 768
Massey, W. E., 722, 774
Matsubara, T., 43, 109, 765, 766
Matsuda, H., 765
Matthiessen, A., 535, 771
Mattis, D. C., 187, 267, 272, 405, 660, 662, 767, 769
Mattock, R. D., 766
Maxwell, E., 627, 773
Mazur, P., 766
Mao, H. K., 315, 768
McFeely, F. R., 772
McMenamin, J. C., 768
McMillan, W. L., 214, 492, 497, 670, 627, 689, 690, 694, 696, 767, 770, 772, 774
Megerle, K., 773
Meirav, U., 771
Mermin, N. D., 412, 769
Meyer, L., 722, 773
Migdal, A. B., 481, 489, 540, 665, 770, 773
Mihara, N., 361, 768
Millis, A., 773
Mills, A. P., 168, 766
Minnhagen, P., 247, 615, 622, 766, 772
Mirlin, D. N., 770
Mittag, L., 768
Mitrovic, B., 671, 672, 674, 773
Miyake, S. J., 446, 448, 770
Montroll, E. W., 765
Mook, H. A., 714, 774, 775
Mooradian, A., 474, 475, 770
Morse, R. W., 642, 773
Mozer, B. L., 60, 765
Myers, H. P., 589, 772

Nagaev, E. L., 460, 770
Nakajima, S., 491, 770
Nesbitt, L. B., 773
Newton, R. G., 204, 767
Ng, K. W., 773
Niklasson, G., 341, 768
Nozieres, P., 300, 313, 354, 603, 611, 613, 621, 714, 715, 719, 731, 732, 768, 771, 772, 774

Ogiwara, A., 622, 772
Ohmura, Y., 622, 772
Ohtaka, K., 610, 621, 772
Onsager, L., 38, 251, 302, 686, 689, 690, 765, 768, 774
Opsal, J. L., 545, 771
Osheroff, D. D., 677, 735, 774
Ott, H. R., 773
Overhauser, A. W., 262, 411, 767, 769
Owen, C. S., 670, 773
Palmer, L. H., 660, 663, 664, 670, 773
Pandey, K. C., 367, 768
Pardee, W. J., 610, 624, 772
Parker, W. H., 653, 773
Parks, R. D., 628, 658, 773
Parr, R. G., 61, 765
Parratt, L. G., 625, 772
Patton, B. R., 734, 774
Pekar, S. L., 444, 770
Pelizzari, C. A., 723, 775
Pelzer, H., 765
Penrose, O., 686, 688, 689, 690, 774
Penn, D. R., 403, 610, 769, 772
Pepper, M., 774
Percus, J. K., 58, 677, 691, 765, 744
Peschel, I., 267, 269, 767
Pethick, C., 714, 732, 733, 734, 773, 774
Pfeiffer, L. N., 774
Phillips, J. C., 771
Pines, D., 300, 313, 714, 719, 731, 732, 768, 770, 774
Pitarke, J. M., 768
Platzman, P. M., 697, 767, 774
Polder, D., 465
Pollak, R. A., 772
Powell, C. J., 333
Prange, R. E., 545, 750, 771, 774
Primakoff, H., 62
Pudalov, V. M., 768
Puff, R. D., 361, 696, 698, 768, 774

Quinn, J. J., 300, 301, 302, 363, 768
Quong, A. A., 768

Rajagopal, A. K., 486, 770
Reatto, L., 694, 774
Reik, H. G., 465, 770
Reissland, J. A., 487, 770
Reynolds, C. A., 627, 773
Reynolds, D. C., 596, 772
Rice, M. J., 729, 774
Rice, T. M., 247, 366, 634, 767, 768, 769, 773
Richardson, R. C., 774
Rickayzen, G., 628, 773
Riffe, D. M., 769, 772
Ritchie, R. H., 371, 768
Robins, J. L., 333
Rodriquez, L. J., 698, 775
Ron, A., 714, 775
Rosen, A., 313, 768
Rosengren, A., 606, 772
Rowell, J. M., 492, 493, 670, 770, 773

Rubio, A., 768
Runge, E., 766
Ruvalds, J., 703, 775

Sachs, R. G., 475, 770
Saint James, D., 771
Samwer, K., 382, 769
Santoro, G. E., 341, 358, 768
Sawada, K., 300, 768
Scalapino, D. J., 664, 670, 770, 771, 773
Schafroth, M. R., 627, 628, 773
Schiff, L. I., 11, 53, 438, 443, 459, 765
Schilling, A., 628, 773
Schmidt, P., 767
Schneider, T., 355, 768
Schönhammer, K., 214, 391, 398, 614, 767, 769, 772
Schotte, K. D., 461, 619, 770, 772
Schotte, U., 619, 772
Schrieffer, J. R., 47, 389, 480, 482, 515, 516, 539, 566, 627, 628, 728, 766, 769, 770, 771, 773
Schweber, S. S., 187, 767
Schwinger, J., 95, 766
Sears, V. F., 694, 698, 699, 775
Segall, B., 518, 771
Seitz, F., 300, 301, 768
Serene, J. W., 668, 670, 769, 773
Sernelius, B. E., 364, 768
Serin, B., 773
Sham, L. J., 337, 487, 768, 769, 770
Shaw, W., 663, 670, 773
Shekhtman, L., 652, 773
Sherrington, D., 687, 774
Shirley, D. A., 772
Sholl, C. A., 312, 768
Shtrikman, H., 771
Shukla, R. C., 487, 770
Shung, K. W., 364, 768
Sierro, J., 402, 769
Sigel, J. L., 536, 771
Silverstein, S. D., 387, 769
Simanek, E., 582, 772
Simpson, J. A., 768
Singh, R. N., 47, 770
Singwi, K. S., 338, 347, 348, 349, 355, 356,
Sixtensson, P., 589, 772
Sjolander, A., 338, 768
Skold, K., 723, 775
Smith, H., 774
Smith, J. L., 773
Smith, N. V., 589, 590, 769, 772
Smodyrev, M. A., 440, 770
Sofo, J. O., 402, 769
Solna, J., 775
Sondheimer, E. H., 519, 766, 771
Srivastava, P. L., 487, 779
Staver, I., 486, 770
Stephen, M. J., 768
Stewart, G. R., 390, 769
Stirling, W. G., 703, 775
Stormer, H. L., 748, 775
Sturm, K., 590, 771
Sueoka, O., 372, 768
Sunjic, M., 603, 624, 626, 772
Svensson, E. C., 694, 698, 775
Swan, J. B., 333
Swanson, N., 768
Swihart, J. C., 487, 491, 663, 670, 770, 773
Sykes, J., 729, 773

Takegahara, K., 544, 545, 771
Tanabe, Y., 610, 621, 772
Taylor, B. N., 773
Taylor, P. L., 179, 766
Taylor, R., 330, 487, 548, 549, 768, 770, 771
Teller, E., 475, 770
Tenn, J. S., 696, 698, 774
Thaler, B. J., 771
Thomas, D. G., 32, 766, 767
Thomas, H., 768
Thomas, L. H., 323, 768
Thouless, D. J., 761, 774
Tiablikov, S., 455, 770
Ting, C. S., 344, 768
Tinkham, M., 660, 663, 664, 670, 773
Tischler, J. Z., 768
Toda, M., 771
Tomlinson, P. G., 494, 770
Tomonaga, S., 153, 256, 257, 766, 767
Tosi, M. P., 338, 768
Tracy, J. C., 369, 768
Tremblay, A. M. S., 428, 429, 769
Tsui, D. C., 480, 742, 750, 770, 775

Valls, O. T., 729, 774
Van Der Lage, F. C., 666, 773
Van Haeringen. W., 245, 246, 767
Van Harlingen, D. J., 773
Varga, B. B., 276, 767
Vashishta, P., 338, 355, 341, 768
Vidberg, H. J., 668, 670, 773
Vignale, G., 341, 346, 358, 768
Vilk, Y. M., 413, 428, 429, 769
Vollhardt, D., 715, 775
Vosko, S. H., 347, 365, 366, 767

Wada, Y., 770
Wagner, H., 412, 769
Walecka, J. D., 766
Wang, S., 544, 545, 771
Wannier, G. H., 594, 772
Ward, J. C., 513, 771
Watabe, M., 491, 582, 770, 772
Weiss, G. H., 765
Werthamer, N. R., 635, 735, 736, 740, 772, 773
Wertheim, G. W., 370, 605, 769, 772
West, K. W., 774
Wheatley, J. C., 734, 775
Wiegmann, P., 387, 769
Wigner, E. P., 17, 40, 157, 300, 307, 311, 313, 313, 434, 765, 766, 768, 770
Wilkins, J. W., 37, 38, 391, 392, 393, 491, 769, 770, 771, 773, 774
Williams, F. E., 232, 767
Winzer, K., 382, 769
Wolf, E. L., 773
Wolfe, P., 715, 775
Wolff, P. A., 47, 389, 766, 769
Woll, E. J., 487, 770
Wollman, D. A., 659, 660, 773
Woo, C. W., 722, 774
Woods, A. D. B., 698, 699, 702, 773, 775
Wright, G. B., 474, 475, 770
Wright, W. H., 773
Wu, F. Y., 23, 765
Wu, J. W., 592, 772

Yamashita, J., 455, 770
Yang, C. N., 686, 775
Yennie, D. R., 750, 775
Yosida, K., 383, 769

Zaringhalam, A., 734, 774
Zawadowski, A., 561, 703, 771, 775
Zeller, H. R., 567, 771
Zener, C., 46, 766
Zhao, Y., 770
Zienau, S., 765
Ziman, J. M., 487, 499, 500, 537, 770
Zschack, P., 768
Zubarev, D. N., 694, 773
Zwicknagl, G., 401, 769

Subject Index

Advanced Green's function: *see* Green's functions
Allen formula, 591–592
Alpha-squared-F
  regular, 492, 527–533, 542, 668
  transport, 527–533, 542, 548
Anderson model, 44, 46–48, 389–403, 569
Anharmonic interactions, 10–11, 708
Antiferromagnetism, 411, 416–422
Anti-time ordering, 96, 154
Auger decay, 605, 624

Ballistic transport, 575
Bardeen–Cooper–Schrieffer: *see* BCS
BCS theory, 386, 420, 480, 627–644, 678, 713
Bethe ansatz, 375
Bethe lattice, 247–256, 292–293
Bijl–Feynman formula, 703–707
Bipartite, 418
Bogoliubov theory, 419, 427
Boltzmann equation, 499–505
Born–Oppenheimer approximation, 27
Bose–Einstein condensation, 627, 677
Brillouin–Wigner perturbation theory, 434–437, 489

Cayley tree: *see* Bethe lattice
CBF: *see* Correlated basis functions
CDW: *see* Charge-density wave
Charge-density waves, 411
Charge–spin separation, 406
Charging energy, 301, 568
Collision term
  impurity scattering, 501
  electron–phonon scattering, 525
Commutation relations
  particles, 13, 118
  phonons, 3
  photons, 56
  spins, 37–39
  Tomonaga model, 258
Complex time, 110–112
Compressibility, 352
Compressibility sum rule, 353
Compton scattering, 366
Condensate, 682, 697
Conductivity
  impurity scattering, 499–514
  polarons, 517–523
Configurational coordinates, 232, 235–236, 465–466
Conserving approximation, 337
Cooper instability, 628–632
Coordination number, 20, 43, 251, 403, 412, 455
Correlated basis functions, 690–696, 721
Correlation energy, 295, 300
Correlation function
  current–current, 165, 182, 229, 463, 582, 596
  density–density, 321, 536, 726
  force–force, 535–537, 579–585
  magnetic susceptibility, 175, 341–342, 356–358, 727
  two-particle, 167
Coulomb blockade, 569–575
Coupling constant integration, 148, 350
Creation operator
  condensate, 682
  electrons, 97
  particles, 13
  phonons, 2, 99, 222
  photons, 57
  spins, 42
  Tomonaga model, 258
Cumulant expansion, 142–154, 241–247, 319, 448–454, 614–619
Current
  electrical, 23, 161
  energy, 25, 178
  heat, 26, 179, 546
  particle, 24, 178
  polarization, 23
Current–current correlation function: *see* Correlation function

Dangerous diagrams, 151
Debye–Waller factor, 461
Deformation constant, 32
Deformation potential, 31

Density matrix, 60, 687
Density of states
  Bethe lattice, 251
  free electrons, 189
Density operator, 17, 23
Destruction operator: *see* Creation operator
Diagonal transitions, 457
Dielectric constant, 33
Dielectric function, 106, 174, 277, 287, 317
  definition, 321
  electrons and phonons, 468–474
  Hubbard, 336
  impurity scattering, 536
  Lindhard: *see* Random phase approximation
  Lorenz–Lorentz, 287, 336
  random phase approximation, 306
  Singwi–Sjolander, 338
  Thomas–Fermi, 323
Dipole approximation, 285
Disconnected diagrams, 83
Disentangling of operators, 221–224
Distribution function
  classical, 500, 49
  Wigner, 157–160, 549
Dressed interaction, 725
Drude formula, 580, 589
Dyson's equation, 86–90, 99–102, 105, 128–135, 155–157, 200, 215

Edge singularities: *see* X-ray edge singularity
Effective mass, 128, 368, 435, 491
Einstein model for phonons, 224
Electrical current operator, 23
Electron–electron interactions, 19
Electron–phonon interactions, 26–36
  deformation potential, 31, 227
  mass enhancement, 491–495, 544–545, 627
  piezoelectric, 32, 227
  polar, 34
Electron self-energy
  correlation, 300
  exchange, 92, 298
  from phonons, 87, 134
Electronic polaron, 401
Elliott formula, 595–600
Emission spectra, 234
Energy current, tight-binding, 24
Entropy, 142, 177–179, 185
Equal spin pairing, 741
Equations of motion, 327, 339, 636
Exchange, 295–300
  helium, 679–682
  Hubbard model, 409–411
  screened, 306, 363
Excitonic insulator, 634
Excitons
  metals, 607–611
  Wannier, 592–600, 632
Exponential resummation, 452

f-sum rule, 290
  longitudinal, 359
  transverse, 580
Fano–Anderson model, 44, 208–217, 247, 292
Fermi liquid theory, 713–720
Fermi wave vector, 74, 196, 257, 296, 305
Ferromagnetism, 411, 413–416
Feynman diagrams, 81, 94, 565
Feynman theorem, 221
Filling factor, 745
Forward scattering amplitude, 725
FQHE: *see* Fractional quantum Hall effect
Fractional quantum Hall effect, 750–761
Frequency summations, 136–139
Friedel oscillations, 196–197, 346
Friedel sum rule, 196, 213, 317, 622
Frustration, 418
Fumi's theorem, 198, 204, 292, 384, 614

Gauges, 49
  Coulomb, 50
  Lorentz, 52
  magnetic fields, 743
  transverse, 50
Grand canonical ensemble, 112
Gaussian distribution, 227
Gradient expansion, 554–556
Green's functions
  advanced, 96, 118–128
  anti-time-ordered, 96
  cumulant expansion, 152–154, 241–247, 448–454, 619–621
  electron, 71, 12–115
  Matsubara, 112
  phonons, 75, 115–117
  photons, 102–106, 117, 276
  retarded, 96, 118–128, 159
  time-ordered, 71–76, 96
  unperturbed, 73
Ground state energy
  homogeneous electron gas, 295–311, 319–320, 344, 350
  impurities, 204–207
  thermodynamic potential, 148–150
GW approximation, 365

H_{sd}, 46, 376, 383
Hall effect, 745–747
Hamiltonian
  Anderson model, 46
  bare phonon, 483
  electron–phonon, 27–36
  Fano–Anderson model, 44, 207
  Frohlich, 34–36, 433–444, 517–523
  harmonic chain, 6
  harmonic oscillator, 1
  Heisenberg model, 38
  helium, 678
  homogeneous electron gas, 22, 295
  Hubbard model, 22

Hamiltonian (*cont.*)
Ising model, 38
Kondo model, 46, 375
lattice gas model, 42
phonons, 10
particles, 12–19
photons, 55
self-consistent field, 326
spin, 36–48
tight-binding model, 20
tunneling, 561
XY model, 3
Harmonic oscillator, 1–11
in electric field, 5–6, 30–31
Hartree interaction, 297
Hartree–Fock approximation, 283, 298, 679–682
Heat current, 26
Heavy fermions, 390, 401, 635
Heisenberg Hamiltonian, 38
Hole
exchange and correlation, 310
core hole, 600, 605
in Fermi sea, 72, 334, 605
in filled band, 168, 412
valence band, 594
Holstein formula, 537, 543–544
Homogeneous electron gas, 19
Hubbard model
definition, 22, 571
magnetic phases, 411–429
paramagnons, 729
solution, 403–429

Impurity scattering, 199–207, 499–514, 558–561, 584
Incomplete phonon relaxation, 606, 623
Independent boson model, 31, 218–247, 372, 441, 452, 619
Infrared divergence, 270, 620
Interactions
anharmonic, 10–11
electron–electron, 19
electron–phonon, 26–36
particle–particle, 15–19
Interband absorption, 588–590
Intermediate coupling, 438
Itinerant magnetism, 375
Ising model, 38, 251–256
Isotope effect, 627, 632

Jahn–Teller effect, 26
Jellium: *see* Homogeneous electron gas

Kohn anomaly, 487
Kondo
effect, 376, 621
model, 46–48, 375–389
resonance, 388, 397, 400
temperature, 377, 386–389
Kramers–Kronig relation, 360, 363, 479, 497
Kubo formula, 163–165, 234, 505, 537, 562

Ladder diagrams, 510, 540, 598
Lagrangian
particles, 12
photons, 53
Landau levels, 742
Laughlin wave function, 755
Lennard-Jones potential, 678, 694
Lattice gas model, 42
Lehmann representation, 122, 139, 397, 507
Levinson's theorem, 196
Linear response, 165
Linked cluster expansion: *see* Cumulant expansion
Local field corrections, 336, 427
Lowering operator
phonons, 2
particle, 14–15
spins, 37
Luttinger liquid, 276
Luttinger model, 264–276

Magnetic impurities, 375
Magnetic susceptibility, 174–177, 264
Magneto rotons, 757–759
Mahan singularity: *see* X-ray edge divergence
Mass operator, 90
Matthiessen's rule, 535
Mean field theory, 642
Mean free path, 369
Meissner effect, 627
Metallic hydrogen, 315–316
Mixed valence, 391
MND, for Mahan, Nozieres and deDominicis, 603. 621–623
Mobility
acoustical phonons, 524–534
CdTe, 518
polarons, 517–523
Moments
first, 226
second, 292
third, 361

Nesting, 411, 418
Noncrossing approximation, 391, 397
Nondiagonal transitions, 457
Non-Fermi liquid behavior, 276
No-phonon transition, 232, 461
Number operator, 442, 620
Nyqvist theorem, 535

Occupation number
bosons, 14
electrons, 75
phonons, 8
Off-diagonal long range order, 686–690
Onsager relations, 177–182
Optical properties, 579–580
Optical theorem, 194
Orthogonality catastrophe, 612–621

Pair distribution function, 59–62, 308–311, 341, 346, 688, 754
Paramagnetic susceptibility, 174, 264, 356
Paramagnon theory, 714, 728
Particle current, 24
Particle–hole symmetry, 412
Partition function, 21
Perturbation theory
  Brillouin–Wigner, 434
  on the mass shell, 438
  Rayleigh–Schrödinger, 438
Phase shift, 191–199
Phonons, 1–11
  anharmonic interactions, 10–11
  cloud, 442
  drag, 382
  first quantization, 1
  force between electrons, 1, 31
  Kohn anomaly, 47
  metals, 482–487
  number, 442
Photons, 48–57, 276–291
Piezoelectric interaction, 32, 471
Plasma frequency
  ions, 280, 468, 483
  metals, 332–333
  semiconductors, 472–475
  tight-binding, 406–408
Poincare's theorem, 211
Polar interaction, 34
Polariton, 276–291
Polarizability, 284
Polarization operator, 90, 284
Polaron
  electronic, 401, 463
  Frohlich, 433–444
  linked-cluster, 448–454
  optical absorption, 585–588
  p-wave states, 44
  small, 454–463
  strong-coupling, 444–448
Pseudopotential, 19, 31, 197

QBE: *see* Quantum Boltzmann equation
QD: *see* Quantum dot
QHE: *see* Quantum Hall effect
Quantum Boltzmann equation, 95, 158, 549–561
Quantum conductance, 575
Quantum dots, 561–571
Quantum Hall effect, 742, 747–750
Quantum lattice gas, 43
Quasiparticles, 365

Raising operator
  phonons, 2
  particles, 14–15
  spin, 37
Raman scattering, 278–279, 474–475, 495
Random phase approximation, 306, 325, 406
Rayleigh–Schrödinger perturbation theory, 300, 434, 438
Reaction matrix, 189–192
Relaxation energy, 30
Relaxation time
  acoustical phonons, 524–534
  approximation, 500
  degenerate semiconductors, 481
  energy, 526–531
  impurity scattering, 502
  metals, 484
  polarons, 444
  temperature, 531–534
  transport, 500
Renormalization castrophe, 607
Renormalization factor, 127, 216, 365
Representations
  Heisenberg, 66
  interaction, 67
  Schrödinger, 66
Resistivity, 505, 536–537, 558–561
Retarded Green's function: *see* Green's functions
Rotons, 703
  bound state, 703
  ^{4}He, 702–703
RPA: *see* Random phase approximation
r_s, 296, 467
Rules for constructing diagrams, 90–92, 135

Scattering resonance, 208
Schrieffer–Wolff transformation, 47
Screening charge density, 317, 346
Seebeck coefficient: *see* Thermopower
Seitz's theorem, 301, 353
Self-consistent field, 325
Self-energy
  Dyson's equation, 86–90, 132
  electron
    electron–electron interaction, 93–94, 295–311
    electron–phonon interaction, 134, 475–481, 487–495
  phonon, 90, 92–3, 482–487
s-d Interaction: *see* H_{sd}
SDW: *see* Spin-density waves
Single mode approximation, 362, 758
Slater determinant, 61, 612
Slave bosons, 391
S matrix
  impurity averaging, 199–200
  time-loop, 95–102
  time-ordered, 70–76
  tau-ordered, 130
s-p approximation, 732
Spectral function
  dielectric function, 322
  electron, 102, 122–128, 212
  phonon, 122, 125

Spin sound, 727
Spin waves, 262
Spin-density waves, 411, 418
Static form factor, 57
Static structure factor, 57, 321, 338, 703, 722, 754
Stoner criteria, 407, 414, 422
Strong-coupling theory
  polarons, 444–448
  superconductors, 639, 663–670
Structure factor: *see* Static structure factor
Sudden approximation, 367, 700–702
Sudden switching, 236–241
Sum rules, 123, 212, 225, 358, 428–429
Superconductors
  anisotropic s-wave, 644
  coherence factors, 639
  density of states, 643
  dirty, 662
  d-wave pairing, 643–644
  energy gap, 639, 642
  excitation energy, 639
  high-temperature, 628
  infrared absorption, 660–664
  p-wave, 635, 643, 659
  renormalization coefficient, 667, 671
  strong coupling, 639, 646, 663–670
  transition temperature, 627, 670–674
  triplet pairing, 635, 736–740
  weak, 627

T-matrix, 192, 503
Tamm–Dancoff approximation, 437
Tau ordering, 112
Thermal conductivity, 180
Thermopower, 181, 401–402, 546–549
Tight-binding model
  definition, 20
  electrical current, 23
  energy current, 25
  interactions, 22
  particle current, 23
  small polarons, 454
Time ordering operator, 68, 96, 154
TiO_2, 463, 465, 466
Tomonaga model, 256–276
Trace, 109, 116, 120, 130, 170, 220
Transport
  ^{3}He, 729, 734
  impurity scattering, 499, 514
  insulators, 517–523
  metals, 527–528
  semiconductors, 528–531
Triplet pairing, 734
Tunneling
  conductance, 646
  Josephson, 652–660
  normal metals, 561–567
  quantum dots, 571–575
  superconductors, 644–660
Two fluid model, 703, 710–713

Ultrasonic attenuation, 642
Unitary superfluid, 742

Vacuum polarization, 83, 132
Van der Monde determinant, 754
Van Hove singularity, 415
Vertex corrections, 343, 509, 538–540, 607–609
V_{sd}, 46, 376, 383

Ward identity, 513–517, 538–539
WDF: *see* Wigner distribution function
Wick's theorem, 76, 100, 131, 391
Wigner lattice, 311–316
Wigner distribution function, 157–160, 338, 549–561, 689
Wigner–Seitz model, 311–312

XPS, for X-ray photoelectron spectroscopy, 603, 624–626
X-ray edge divergence, 403, 603, 607
XY model, 38

Yukawa potential, 325

Zero phonon line, 232, 461
Zero sound, 723
Ziman formula, 537