Principles of Quantum Condensed Matter Physics
Part III Major Option
Michaelmas 2002

P B Littlewood

May 15, 2003

Contents

Chapter 1

Introduction

1.1 Theories and models in condensed matter physics

Solid state physics is concerned with the abundance of properties that arise when atoms are amalgamated together. Much of what we think of as "core physics" is deliberately reductionist; we look for the very simplest unified description of a basic phenomenon, and the progress of much of basic physics has always been a progress toward grander unified theories, each of which is simpler (at least in concept) than the previous generation.

Condensed matter physics is not like this. The Hamiltonian is not in doubt - it is the Schrödinger equation for the many particle system:

$$H_{elec} = -\sum_i \frac{\hbar^2}{2m}\nabla_i^2 + \sum_I \frac{P_I^2}{2M_I} + \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2}\sum_{i\neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2}\sum_{I\neq J} \frac{Z_I Z_J e^2}{|\mathbf{R}_i - \mathbf{R}_j|} , \tag{1.1}$$

where the $\mathbf{r}_i, \mathbf{R}_I$ label the coordinates of the electrons and the ions respectively, Z_I, M_I are the nuclear charge and mass. The terms in Eq. (1.1) represent, in order, the kinetic energy of the electrons, the kinetic energy of the nuclei, and the Coulomb interaction between electron and nucleus, electron and electron, and between nucleus and nucleus. In some sense, a complete theory of solids would be to solve the Schrodinger equation and then apply all the standard methods of statistical physics to determine thermodynamic and physical properties. From this point of view, there is no "fundamental" theory to be done, although the calculations may indeed be complex (and in fact, impossible to perform exactly for solids with macroscopic numbers of atoms). Because an exact solution for a macroscopic number of atoms is impossible, we have to treat Eq. (1.1) by a sequence of approximations (for example, perhaps fixing the ions in place, or neglecting electron-electron interactions) that will make the problem tractable.

This view of condensed matter physics as a series of approximations that is widely held and severely incomplete. Suppose for a moment that we could solve the full Hamiltonian, and we would then have a wavefunction describing some 10^{23} particles that contained all of the physics of solids. Writing the solution down would be hard enough, but comprehending its meaning would be beyond us. Condensed matter physics is about phenomena, from the mundane (why is glass transparent), to the exotic (why does ^{3}He become a superfluid). There are a host of physical phenomena to be understood, and their explanation must involve more than just detailed calculation.

Understanding a phenomenon involves building the simplest possible model that explains it, but the models are more than just approximations to Eq. (1.1). Models, and the theories which they give rise to, elucidate paradigms and develop concepts that are obscured by the complexity of the full Hamiltonian. The surprise about condensed matter physics is that there are so many *different* theories that can arise from such an unprepossessing Hamiltonian as Eq. (1.1).

1.1.1 "The Properties of Matter"

A venerable route to condensed matter physics, and one followed by almost all textbooks, is to find ways of making approximate calculations based on the full Schrödinger equation for the solid. Making approximate, but quantitative calculations of the physical properties of solids has been one of the enduring agendas of condensed matter physics and the methods have acquired increasing sophistication over the years. We would like to understand the cohesion of solids – why it is, for example that mercury is a liquid at room temperature, while tungsten is refractory. We wish to understand electrical and optical properties – why graphite is a soft semi-metal but diamond a hard insulator, and why GaAs is suitable for making a semiconductor laser, but Si is not. Why is it that some materials are ferromagnetic, and indeed why is it that transition metals are often magnetic but simple s-p bonded metals never? We would like to understand chemical trends in different classes of materials – how properties vary smoothly (or not) across the periodic table. These, and many other physical properties we now know how to calculate with considerable accuracy by sophisticated computational techniques, but more importantly (and especially for the purposes of this course) we can understand the behaviour straightforwardly, and describe the physical properties in a natural fashion.

To get this understanding we need to develop the basic machinery of the quantum mechanics of periodic structures, especially the concept of electronic bandstructure describing the dispersion relation between the electron's energy and momentum. We also need to understand how the largest effects of interactions between electrons can be subsumed into averaged effective interactions between independent quasiparticles and the background medium. A large part (and certainly the initial part) of this course will be to set up this fundamental machinery.

This is a tidy scheme, but it will get us only part way to the goal. It will generate for us a landscape upon which we can build new models and new theories.

1.1.2 Collective phenomena

There is another view of condensed matter physics which we shall also explore, that is less concerned with calculation and more concerned with phenomena per se. The distinguishing character of solid state systems is that they exhibit *collective* phenomena, that are properties of macroscopic systems and that exist only on account of the many-degree-of-freedom nature of the system.

A familiar example is a phase transition (between liquid and solid, say) which is a concept that can only apply to a macroscopic ensemble. Condensed matter systems have collective modes that are a consequence of their order; both a solid and a liquid support longitudinal sound waves, but a solid that has a nonzero shear stiffness has also transverse sound modes. In fact the existence of shear waves we might choose to define as the characteristic feature distinguishing a solid from a liquid or gas. We can say that solidity is a *broken symmetry* (with the symmetry being broken that of translational invariance); because of the broken symmetry, there is a new collective mode (the shear wave). Because of quantum mechanics, the waves are necessarily quantised as phonons, and they are a true quantum particle, with Bose statistics, that interact with each other (due to anharmonicity) and also with other excitations in the solid. This idea, that a broken symmetry can generate new particles, is one of the central notions of condensed matter physics – and of course of particle physics too.

A different example is the behaviour of electrons in a semiconductor. If one adds an electron into the conduction band of a semiconductor it behaves like a particle of charge $-|e|$, but a mass different from the free electron mass due to the interaction with the lattice of positively charge ions as well as all the other electrons in the solid. But we know that if we remove an electron from the valence band of the semiconductor, it acts as a *hole* of charge $+|e|$; the hole is in fact a collective excitation of the remaining 10^{23} or so electrons in the valence band, but it is a much more convenient and accurate description to think of it as a new fermionic *quasi*-particle as an excitation about the ground state of the solid. The electrons and holes, being oppositely charged, can even bind together to form an exciton - the analog of the hydrogen atom (or more directly positronium), which however has a binding energy considerably reduced from hydrogen, because the Coulomb interaction is screened by the dielectric constant of the solid, and the electron and hole masses are different from the electron and proton in free space.

The solid is a new "vacuum", inhabited by quantum particles with properties which may be renormalised from those in free space (e.g. photons, electrons) or may be entirely new, as in the case of phonons, plasmons (longitudinal charge oscillations), magnons (waves of spin excitation in a magnet), etc. In contrast to the physical vacuum, there are different classes of condensed matter systems which have different kinds of vacua, and different kinds of excitations. Many of these new excitations arise because of some "broken" symmetry , for example, magnetism implies the existence of spin waves, and solidity implies the existence of shear waves. Some of these phenomena – superconductivity, superfluidity, and the quantum Hall effect come to mind – are remarkable and hardly intuitive. They were discovered by experiment; it seems unlikely that they would ever have been uncovered by an exercise of pure cerebration starting with the Schrodinger equation for 10^{20} particles.

Solid state systems consist of a hierarchy of processes, moving from high energy to low; on the scale of electron volts per atom are determined the cohesive energy of the solid, (usually) the crystal structure, whether the material is transparent or not to visible light, whether the electrons are (locally) magnetically polarised, and so on. But after this basic landscape is determined, many further phenomena develop on energy scales measured in meV that correspond to thermal energies at room temperature and below. The energy scales that determine magnetism, superconductivity, etc. are usually several orders of magnitude smaller than cohesive energies, and the accuracy required of an *ab initio* calculation would be prohibitive to explain them. Although all condensed matter phenomena are undoubtedly to be found within the Schrödinger equation, they are not transparently derived from it, and it is of course better to start with specific models that incorporate the key physics; we shall see many of them. These models will usually be simply of interactions between excitations of the solid, with sets of parameters to describe them – parameters which are usually estimated, or derived from experiment.

1.2 Outline of the course.

This course breaks up into several sections that have different goals, but largely interrelated material.

In the first chapters, we will discuss the landscape of condensed matter physics, and introduce the basic tools of band theory applied to periodic solids. Much of this will be done within the "independent particle" approximation applied to the electrons, but we will also address the limitations of this approach, as well as its successes. However, the fundamental electronic structure of solids is the basis on which everything else is constructed.

We will then shift perspective to take a view of solids in terms of their collective behaviour. Nearly all of the measurements that one can make on a system can be viewed as measuring the response of the solid to an external perturbation – for example, with light one measures the dielectric response to an applied oscillating electric field. When the perturbation is weak, this behaviour can be formulated in terms of a linear response function, and this linear response function also tells us about collective modes of oscillation.

Lastly, we shall survey just a few of the occasionally surprising collective phenomena that occur in condensed matter systems.

1.3 Books

There are many good books on solid state and condensed matter physics, but the subject is rich and diverse enough that each of these contains both much more and much less than the topics covered in this course. The two classic textbooks are Kittel, and Ashcroft and Mermin. These are both at the correct level of the course, and have the virtue of clear exposition, many examples, and lots of experimental data. Slightly more concise, and a little more formal in places is Ziman. Grosso and Parravicini has a somewhat wider coverage of material, but much of it goes well beyond the level of detail required for this course. Marder is at about the right level (though again with more detail than we shall need), and has a nice blend of quantum properties with statistical and classical properties.

- C.Kittel, *Introduction to Solid State Physics, 7th edition*, Wiley, NY, 1996 .
- N.W.Ashcroft and N.D.Mermin, *Solid State Physics*, Holt-Saunders International Editions, 1976.
- J.M.Ziman, *Principles of the Theory of Solids*, CUP, Cambridge, 1972.
- M.P. Marder, *Condensed Matter Physics*, Wiley, NY, 2000.
- G.Grosso and G.P.Parravicini, *Solid State Physics*, AP, NY, 2000.
- A very good book, though with a focus on statistical and "soft" condensed matter that makes it not so relevant for this course, is P.M.Chaikin and T.Lubensky, *Principles of Condensed Matter Physics*, CUP, Cambridge, 1995.

Some more specialised texts, all going well beyond the needs of this course:

- On superconductivity:
  Tinkham, *Introduction to Superconductivity*, McGraw Hill, 1975.

- On metal-insulator transitions and strongly correlated systems in general - from an experimental viewpoint not too overburdened with theory:
  N. Mott, *Metal-Insulator Transitions*, Taylor and Francis, London, 1990.

- On semiconductor structure and bonding:
  J.C.Phillips, *Bonds and Bands in Semiconductors*, AP, NY 1973.

- Two books with more formal discussions of the many-body problem, for those of a theoretical inclination, and not for this course:
  S.Doniach and E.H. Sondheimer, *Green's functions for Solid State Physicists*, 2nd Edition, Imperial College Press, 1998.
  G.Rickayzen, *Green's functions and condensed matter*, AP, NY, 1980.

1.4 These notes

These notes are designed as a complement to the lectures, and as a complement to reference books. They are not to be memorised for examination: often they include detailed derivations that are there to satisfy the curious, for completeness, and for background. The lectures will be presented using more qualitative and physical descriptions.

In a few places, and particularly where I shall be discussing material that is not easy to find collected in textbooks, the notes are much more lengthy. Material which is explicitly non-examinable is placed in small type; but in general, no detailed derivations will be required for examination. You may find it worthwhile, however, to work through some of this at least once.

Appendix A collects mathematical techniques that are useful for this course. Most of this course is in Fourier space, so you need to be comfortable with Fourier series and Fourier transforms (section A.1); the calculus of variations (A.2) and the elementary theory of analytic functions (A.3) are also used in some of the formal development, though aside from Eq. (A.58) there is nothing here that you need to understand in any depth.

Appendix B (on second quantisation) is included for those of a theoretical inclination as a primer on many body theory. It is the preferred formalism for doing detailed calculations, and for visualising the physics, though I will not be using it in lecturing (except perhaps as shorthand, on occasion). It is *not* for examination. You should recognise that the *use* of second quantised notation is ubiquitous, even by those who do not use it to perform calculations.

Problems are placed at the ends of the chapters. They vary from the straightforward to the complex, and especially mathematical ones are given a warning asterisk. Problems of such type will not appear in the examination.

Chapter 2

Theory of the electronic structure of solids

2.1 Independent particles in a periodic potential

2.1.1 Periodic structures

An ideal crystal is constructed from the infinite repetitition of identical structural units in space. The repeating structure is called the *lattice*, and the group of atoms which is repeated is called the *basis*. The basis may be as simple as a single atom, or as complicated as a polymer or protein molecule. This section discusses briefly some important definitions and concepts. For a more complete description with examples, see any of the textbooks recommended in the introduction.

Lattice symmetries

The lattice is defined by three fundamental (called *primitive*) translation vectors $\mathbf{a}_i$, $i = 1, 2, 3$. The atomic arrangement looks the same from equivalent points in the unit cell:

$$\mathbf{r}' = \mathbf{r} + \sum_i n_i \mathbf{a}_i \quad \forall \text{ integer } n_i \; . \tag{2.1}$$

The primitive unit cell is the parallelipiped formed by the primitive translation vectors $\mathbf{a}_i$, and an arbitrary lattice translation operation can be written as

$$\mathbf{T} = \sum_i n_i \mathbf{a}_i \tag{2.2}$$

There are many ways of choosing a primitive unit cell, but the lattice so formed is called a *Bravais* lattice.

Often, the most convenient primitive unit cell to use is the *Wigner-Seitz* cell, which is constructed as follows: Draw lines to connect a given lattice

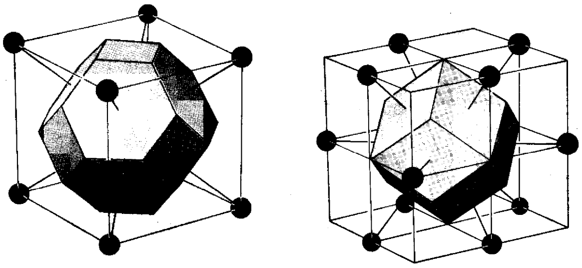


Figure 2.1: . The Wigner-Seitz cell for the BCC and FCC lattices

point to all of its near neighbours. Then draw planes normal to each of these lines from the midpoints of the lines. The smallest volume enclosed in this way is the Wigner-Seitz primitive unit cell.

The are other symmetry operations that can be performed on a lattice, for example rotations and reflections. We call the collection of symmetry operations, which applied about a lattice point, map the lattice onto itself the lattice *point group*. This includes reflections and rotations; for example a 2D square lattice is invariant under reflections about the x and y axes, as well as through axes at an angle of $\pi/4$ to the x and y axes, and rotations through any multiple of $\pi/2$. Remember that adding a basis to a primitive lattice may destroy some of the point group symmetry operations. There are five distinct lattice types in two dimensions, and 14 in three dimensions.

The translational symmetry and the point group symmetries are subgroups of the full symmetry of the lattice which is described by the *space* group. Every operation in the space group consists of a rotation, reflection, or inversion followed by a translation. However, the space group is not necessarily just the sum of the translational symmetries and the point symmetries, because there can be space group symmetries that are the sum of a proper rotation and a translation, neither of which are independently symmetries of the lattice.

The number of possible lattices is large. In three dimensions there are 32 distinct point groups, and 230 possible lattices with bases. Two of the important lattices that we shall meet later are the body-centred and face-centred cubic lattices, shown in Fig. 2.1.

Reciprocal lattice

The reciprocal lattice as a concept arises from the theory of the scattering of waves by crystals. The weak scattering of a wave by a localised potential is familiar from elementary quantum mechanics. If we send incoming radiation of wavevector $\mathbf{k}_o$ onto a potential centred at the point $\mathbf{R}$, at large distances the scattered wave take the form of a circular wave. (See figure Fig. 2.2) The total field (here taken as a scalar) is then

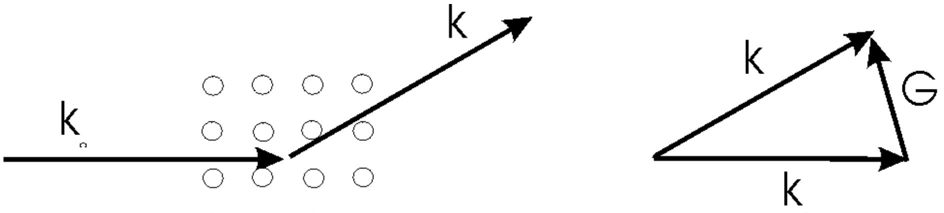


Figure 2.2: Illustration of Bragg scattering from a crystal

$$\psi \propto e^{i\mathbf{k}_o \cdot (\mathbf{r}-\mathbf{R})} + f(\hat{r}) \frac{e^{ik_o|\mathbf{r}-\mathbf{R}|}}{|\mathbf{r}-\mathbf{R}|} \tag{2.3}$$

All the details of the scattering is in the *form factor* $f(\hat{r})$ which is a function of the scattering angle. It depends of course on the arrangement and type of atom in the crystal, as well as the momentum exchanged. For sufficiently large distance from the scatterer, we can write

$$k_o|\mathbf{r}-\mathbf{R}| \approx k_o r - k_o \frac{\mathbf{r}\cdot\mathbf{R}}{r} \tag{2.4}$$

so if we define the scattered wavevector

$$\mathbf{k} = k_o \frac{\mathbf{r}}{r} \tag{2.5}$$

and the momentum transfer

$$\mathbf{q} = \mathbf{k}_o - \mathbf{k} \tag{2.6}$$

we then have for the waveform

$$\psi \propto e^{i\mathbf{k}_o \cdot \mathbf{r}} \left[1 + f(\hat{r}) \frac{e^{i\mathbf{q}\cdot\mathbf{R}}}{r} \right] \quad . \tag{2.7}$$

We must now sum over all the identical sites in the lattice, and the final formula is

$$\psi \propto e^{i\mathbf{k}_o \cdot \mathbf{r}} \left[1 + \sum_i f_i(\hat{r}) \frac{e^{i\mathbf{q}\cdot\mathbf{R}_i}}{r} \right] \quad . \tag{2.8}$$

Away from the forward scattering direction, the incoming beam does not contribute, and we need only look at the summation term. Its behaviour is well known from the theory of Fourier series, and is familiar in one dimension. We are adding together terms with different phases $\mathbf{q}\cdot\mathbf{R}_i$, and these will lead to a cancellation unless the Bragg condition is satisfied

$$\mathbf{q}\cdot\mathbf{R} = 2\pi m \tag{2.9}$$

for all $\mathbf{R}$ in the lattice, and with m an integer (that depends on R). The special values of $\mathbf{q} \equiv \mathbf{G}$ that satisfy this requirement lie on a lattice, which is called the *reciprocal lattice*. We can be sure that they are on a lattice, because if we have found any two vectors that satisfy Eq. (2.9), then their sum also satisfies the Bragg condition.

One can check that the following prescription for the reciprocal lattice will satisfy the Bragg condition. The primitive vectors $\mathbf{b}_i$ of the reciprocal lattice are given by

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a_2} \wedge \mathbf{a_3}}{\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3} \quad \text{and cyclic permutations} \quad . \tag{2.10}$$

The Wigner-Seitz cell of the reciprocal lattice is called the first *Brillouin zone*. This will play an important role in the discussion of electronic states in a periodic potential. Because we have been discussing elastic scattering, we had the two conditions relating incident and outgoing momenta. Conservation of energy requires that the magnitudes of k_o and k are equal, and the Bragg condition requires their difference to be a reciprocal lattice vector $\mathbf{k} - \mathbf{k}_o = \mathbf{G}$. The combination of the two can be rewritten as

$$\mathbf{k} \cdot \frac{\mathbf{G}}{2} = (\frac{G}{2})^2 \quad . \tag{2.11}$$

Eq. (2.11) defines a plane constructed perpendicular to the vector $\mathbf{G}$ and intersecting this vector at its midpoint. The set of all such planes defines those incident wavevectors that satisfy the conditions for diffraction.

2.1.2 Bloch's theorem

We consider the eigenstates ψ of the one-electron Hamiltonian in a periodic potential

$$H\psi(\mathbf{r}) = \left[-\hbar^2\nabla^2/2m + U(\vec{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{2.12}$$

where $U(\mathbf{r}+\mathbf{R}) = U(\mathbf{r})$ for all $\mathbf{R}$ in a Bravais lattice. Bloch's theorem states that they have the form

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}) \tag{2.13}$$

where

$$u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}) \tag{2.14}$$

or, alternatively, that

$$\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n\mathbf{k}}(\mathbf{r}) \tag{2.15}$$

Notice that while the potential is periodic, the wave function consists of *a plane wave* times a periodic function. n is an index, call the *band index*, and we shall see the physical meaning of both n and $\mathbf{k}$ in a moment.

Proof of Bloch's theorem

Here we sketch a proof of Bloch's theorem, and we shall give a somewhat more intuitive (but longer) one later.

First, let us define a *translation operator* $T_{\mathbf{R}}$, which when operating on any function, shifts the argument by a lattice vector $\mathbf{R}$:

$$T_{\mathbf{R}} f(\mathbf{r}) = f(\mathbf{r}+\mathbf{R}) \tag{2.16}$$

It is straightforward[1] to then show that $T_{\mathbf{R}}$ commutes with the Hamiltonian:

$$T_{\mathbf{R}} H = H T_{\mathbf{R}} \tag{2.17}$$

Furthermore

$$T_{\mathbf{R}} T_{\mathbf{R}'} = T_{\mathbf{R}'} T_{\mathbf{R}} = T_{\mathbf{R}+\mathbf{R}'} \quad ; \tag{2.18}$$

the translation operators commute with themselves.

We may now use a fundamental theorem of quantum mechanics; two commuting operators can be chosen to have the same eigenstates, so

$$\begin{aligned} H\psi &= E\psi \\ T_{\mathbf{R}}\psi &= c(\mathbf{R})\psi \end{aligned} \tag{2.19}$$

Applying the results of Eq. (2.17) and Eq. (2.18), we see that the eigenvalues of T must satisfy

$$c(\mathbf{R})c(\mathbf{R}') = c(\mathbf{R}+\mathbf{R}') \tag{2.20}$$

Now let $\mathbf{a}_i$ be three primitive vectors of the lattice, and write

$$c(\mathbf{a}_i) = e^{2\pi i x_i} \tag{2.21}$$

which is just a definition of the x_i, but we have chosen this form because the boundary conditions will in the end force the x_i to be real. Since a general Bravais lattice vector can be written as $\mathbf{R} = n_1\mathbf{a_1} + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, we can then use the rule of Eq. (2.20) to show that

$$c(\mathbf{R}) = c(\mathbf{a_1})^{n_1} c(\mathbf{a_2})^{n_2} c(\mathbf{a_3})^{n_3} = e^{2\pi i(x_1 n_1 + x_2 n_2 + x_3 n_3)} \tag{2.22}$$

which is precisely of the form $c(\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}$ when

$$\mathbf{k} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3 \tag{2.23}$$

and the $\mathbf{b}_i$ are reciprocal lattice vectors that satisfy $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$.

This is precisely Bloch's theorem in the form given in Eq. (2.15).

Boundary conditions and counting states

We are not quite finished, because we have to deal with boundary conditions in an infinite (or at least very large) sample. This will demonstrate that the wavevector $\mathbf{k}$ must be real, and will determine the allowed values of $\mathbf{k}$. We shall use periodic boundary conditions

$$\psi(\mathbf{r} + N_i\mathbf{a}_i) = \psi(\mathbf{r}) \tag{2.24}$$

---

[1] 0perate with the translation operator on $H\psi$ and use the periodic symmetry of the potential

where N_i are integers, with the number of primitive unit cells in the crystal being $N = N_1 N_2 N_3$. Applying Bloch's theorem, we have immediately that

$$e^{iN_i \mathbf{k}\cdot\mathbf{a}_i} = 1, \tag{2.25}$$

so that (see Eq. (2.23)) the general form for the allowed Bloch wavevectors is

$$\mathbf{k} = \sum_i^3 \frac{m_i}{N_i}\mathbf{b}_i, \qquad \text{for } m_i \text{ integral.} \tag{2.26}$$

Thus the volume of allowed k-space per allowed k-point is just

$$\Delta k = \frac{\mathbf{b}_1}{N_1} \cdot \frac{\mathbf{b}_2}{N_2} \wedge \frac{\mathbf{b}_3}{N_3} = \frac{1}{N}\mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3. \tag{2.27}$$

Since $\mathbf{b}_1 \cdot \mathbf{b}_2 \wedge \mathbf{b}_3 = (2\pi)^3 N/V$ is the volume of the unit cell of the reciprocal lattice (V is the volume of the crystal), Eq. (2.27) shows that the number of allowed wavevectors in the primitive unit cell is equal to the number of lattice sites in the crystal. We may thus rewrite Eq. (2.27)

$$\Delta k = \frac{(2\pi)^3}{V} \tag{2.28}$$

The Schrödinger equation in momentum space

We can expand the wavefunction in terms of a set of plane waves that satisfy the periodic boundary conditions:

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.29}$$

whereas the periodic potential $U(\mathbf{r})$ has a plane wave expansion that only contains waves with the periodicity of the reciprocal lattice

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{2.30}$$

where $\mathbf{G}$ are the reciprocal lattice vectors. The momentum components are

$$U_{\mathbf{G}} = \frac{N}{V}\int_{unit\,cell} d\mathbf{r}\, e^{-i\mathbf{G}\cdot\mathbf{r}}\, U(\mathbf{r})\ , \tag{2.31}$$

and since the potential is real $U^*_{\mathbf{G}} = U_{-\mathbf{G}}$.

We now insert Eq. (2.29) and Eq. (2.30) in Eq. (2.12), and obtain, after a little reorganisation of the terms

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \left[\left(\frac{\hbar^2}{2m}k^2 - E \right) c_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \right] = 0 \tag{2.32}$$

Since the plane waves form an orthogonal set, each coefficient in the sum over $\mathbf{k}$ in Eq. (2.32) must vanish, i.e.

$$\left[\left(\frac{\hbar^2}{2m}k^2 - E\right) c_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}}\right] = 0 \tag{2.33}$$

It is often convenient to rewrite $\mathbf{q} = \mathbf{k} - \mathbf{K}$, where $\mathbf{K}$ is a reciprocal lattice vector chosen so that $\mathbf{q}$ lies in the first Brillouin zone, when Eq. (2.33) is just

$$\left[\left(\frac{\hbar^2}{2m}(\mathbf{q}-\mathbf{K})^2 - E\right) c_{\mathbf{q}-\mathbf{K}} + \sum_{\mathbf{G}} U_{\mathbf{G}-\mathbf{K}} c_{\mathbf{q}-\mathbf{G}}\right] = 0 \tag{2.34}$$

We can now see that the wavefunction is of the Bloch form, because Eq. (2.34) mixes plane waves of momentum $\mathbf{q}$ with $\mathbf{q} - \mathbf{G}$, and so

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot\mathbf{r}} u(\mathbf{r}) \ , \tag{2.35}$$

where

$$u(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{-i\mathbf{G}\cdot\mathbf{r}} \tag{2.36}$$

is now a function with the periodicity of the lattice.

Bandstructure

There are in general infinitely many solutions of Eq. (2.34), which is why the energies and wavefunctions are indexed by the band index n as well as the wavevector $\mathbf{k}$. The energy eigenvalue $E_{n\mathbf{k}}$, plotted as a function of $\mathbf{k}$ determines the *bandstructure* – the allowed energy levels. In general, because this is a Hermitian eigenvalue problem , the energy levels will be discretely spaced in the index n. However, because $\mathbf{k}$ is a *parameter* in the eigenvalue equations, the k-dependence of the energy levels is continuous. For this reason, we can talk about energy bands.

The wavevector $\mathbf{k}$ can always be confined to the first Brillouin zone (actually any primitive unit cell, but usually this is the most convenient). Sometimes, it is useful (though redundant) to allow $\mathbf{k}$ to range through all of momentum space, while remembering of course that

$$\begin{aligned} \psi_{n\mathbf{k}+\mathbf{G}}(\mathbf{r}) &= \psi_{n\mathbf{k}}(\mathbf{r}) && (2.37) \\ E_{n\mathbf{k}+\mathbf{G}} &= E_{n\mathbf{k}} && (2.38) \end{aligned}$$

The energy bands themselves have the periodicity of the reciprocal lattice.

Fermi surface and density of states

The ground state of the N-electron system is constructed by occupying all the lowest one-electron levels until the total number of electrons is exhausted. Although we are working in the independent electron approximation, we must still obey Fermi statistics, which will mean that each momentum state can accommodate two electrons of opposite spin (we are assuming for the moment that the energy levels do not depend on spin). To get the counting right, we count only the momentum states within a primitive unit cell – usually the first Brillouin zone.

There are two distinct possibilities. If the number of electrons per unit cell is even, it is possible that a set of bands is completely filled with the higher bands completely empty. In this case we have a band gap – the difference in energy between the highest occupied level and the lowest unoccupied level – and the material is a semiconductor or an insulator as in NaCl or Si. Alternatively, if the bands are partially filled, there will be a surface in momentum space separating the occupied from the unoccupied levels. This *Fermi surface* is at a fixed energy E_F called the *Fermi energy*, and is defined by the momenta that satisfy

$$E_{n\mathbf{k}} = E_F \tag{2.39}$$

If a Fermi surface exists (there may be several branches if the Fermi energy cuts through several bands), the material will be a metal. Clearly, if the number of electrons per unit cell is not even, we must have a metal (Al is an example); but band overlap may yield a metal even if the number of electrons is even (An example is Arsenic, which although it has an odd number of electrons per atom, has two atoms per primitive unit cell, and thus an even number of electrons per cell; two bands overlap near the Fermi surface, and As is an example of a class of materials often called semimetals.)

We often will need to perform summations over k-states, and because the momentum states are very closely spaced, these are best turned into integrals. Typically, we will need quantities such as a weighted average of one electron properties, such as [2]

$$F = \frac{2}{V}\sum_{n\mathbf{k}} F_n(\mathbf{k}) \tag{2.40}$$

Using the result from Eq. (2.28), we turn this into an integral

$$F = 2\sum_n \int_{unitcell} \frac{d\mathbf{k}}{(2\pi)^3} F_n(\mathbf{k}) \tag{2.41}$$

Often, we will be dealing with cases where the function F_n depends on momentum only through the energy $E_n(\mathbf{k})$, and we can then formally introduce

---

[2] The factor of two accounts for spin degeneracy

the density of states in energy $g(E)$, so that

$$F = \int dE\, g(E) F(E) \tag{2.42}$$

The familiar case of dealing with this is for a free-electron-like dispersion (i.e. parabolic)

$$E(k) = \frac{\hbar^2 k^2}{2m} \tag{2.43}$$

We determine the density of states in energy by calculating how many states are enclosed in a thin shell of energy width dE; this is (in 3 dimensions - the answer depends on dimensionality)

$$g(E)dE = \frac{\mathrm{Volume\ of\ shell\ in\ k-space}}{\mathrm{Volume\ of\ k-space\ per\ state}} = \frac{4\pi k^2 dk}{(2\pi)^2/V} \quad , \tag{2.44}$$

hence

$$g(E) = 2\frac{V}{(2\pi)^3} 4\pi k^2 \frac{dk}{dE} = \frac{V}{\pi^2}\frac{m}{\hbar^2}\left(\frac{2mE}{\hbar^2}\right)^{\frac{1}{2}} \quad . \tag{2.45}$$

where the extra factor of 2 appearing accounts for there being two spin states for each allowed k-point. Often, the density of states is given per unit volume, so the favtor of V disappears.

More generally, for any form of $E(k)$, the density of states is

$$g(E) = \sum_n g_n(E) = \sum_n \int \frac{d\mathbf{k}}{4\pi^3} \delta(E - E_n(\mathbf{k})) \ , \tag{2.46}$$

Because of the δ-function in Eq. (2.46), the momentum integral is actually over a surface in k-space S_n which depends on the energy E; $S_n(E_F)$ is the Fermi surface. We can separate the integral in $\mathbf{k}$ into a two-dimensional surface integral along a contour of constant energy, and an integral perpendicular to this surface $dk_\perp$ (see Fig. 2.3). Thus

$$\begin{aligned} g_n(E) &= \int_{S_n(E)} \frac{dS}{4\pi^3} \int dk_\perp(\mathbf{k})\, \delta(E - E_n(\mathbf{k})) \\ &= \int_{S_n(E)} \frac{dS}{4\pi^3} \frac{1}{|\nabla_\perp E_n(\mathbf{k})|}, \end{aligned} \tag{2.47}$$

where $\nabla_\perp E_n(\mathbf{k})$ is the derivative of the energy in the normal direction.[3]

Notice the appearance of the gradient term in the denominator of Eq. (2.47), which must vanish at the edges of the band, and also at saddle points, which exist generically in two and three dimensional bands. Maxima, minima, and saddle points are all generically described by dispersion (measured relative to the critical point) of

$$E(\mathbf{k}) = E_0 \pm \frac{\hbar^2}{2m_x} k_x^2 \pm \frac{\hbar^2}{2m_y} k_y^2 \pm \frac{\hbar^2}{2m_z} k_z^2 \tag{2.48}$$

If all the signs in Eq. (2.48) are positive, this is a band minimum; if all negative, this is a band maximum; when the signs are mixed there is a saddle point. In the vicinity of

---

[3] We are making use of the standard relation $\delta(f(x) - f(x_0)) = \delta(x - x_0)/|f'(x_0)|$

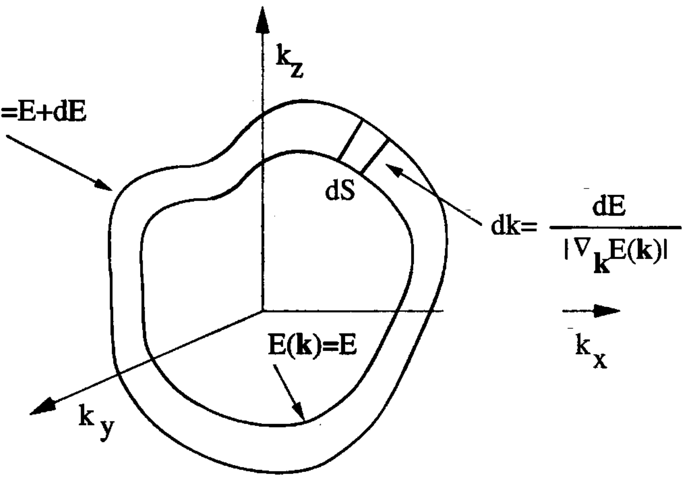


Figure 2.3: Surface of constant energy

each of these critical points, also called van Hove singularities, the density of states (or its derivative) is singular. In two dimensions, a saddle point gives rise to a logarithmically singular density of states, whereas in three dimensions there is a discontinuity in the derivative.

Examples of the generic behaviour of the density of states in one, two and three dimensions are shown in Fig. 2.4.

We now turn to two approximate methods for discussing the bandstructure of solids. In the first, we shall make the assumption that the interaction of the electronic plane waves with the lattice is weak, the so-called nearly-free-electron approximation. In the second we start from the opposite assumption, that the solid is a weakly interacting collection of atoms, and then the best description starts from atomic basis states - this is the tight-binding method.

2.1.3 Nearly free electron description

We have already written down an exact form for the wavefunction in terms of plane waves in Eq. (2.34), which we reproduce below. The general form for the wavefunction of a Bloch state with momentum $\mathbf{k}$ is

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}}, \tag{2.49}$$

where the coefficients $c_{\mathbf{k}}$ are determined by the solution of the set of equations

$$\left[\left(\frac{\hbar^2}{2m}(\mathbf{k}-\mathbf{K})^2 - E\right) c_{\mathbf{k}-\mathbf{K}} + \sum_{\mathbf{G}} U_{\mathbf{G}-\mathbf{K}} c_{\mathbf{k}-\mathbf{G}}\right] = 0\ . \tag{2.50}$$

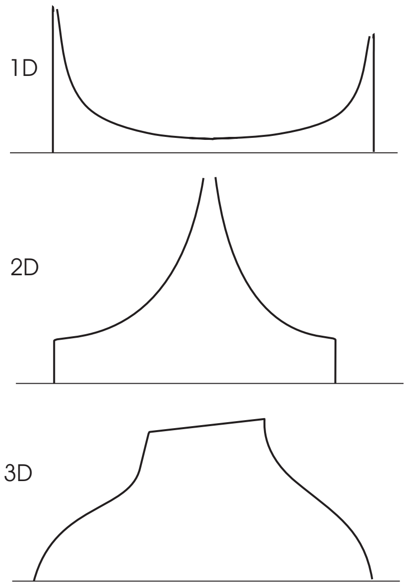


Figure 2.4: Density of states in one (top curve), two (middle curve) and three (lower curve) dimensions

Remember that **k** is restricted to be within the first Brillouin zone.

If the periodic potential vanishes, the only nonzero solutions correspond to the set of energy eigenvalues

$$E = E^0_{\mathbf{q}-\mathbf{K}} \tag{2.51}$$

where we write the kinetic energy of a free particle of wavevector **k** as $E^0_{\mathbf{k}} = \hbar^2 k^2/2m$. Notice that this is not quite the same as the free electron picture, because we still have the reciprocal lattice vectors chosen to index the bands back to the first Brillouin zone; this is indicated in Fig. 2.5. If the potential is weak, we can then incorporate its effects on the energy level spectrum and wavefunctions by perturbation theory. Let's multiply the potential U by a parameter λ in order to keep track of orders of the perturbation theory, viz.

$$\begin{aligned} c_{\mathbf{k}} &= c^{(0)}_{\mathbf{k}} + \lambda c^{(1)}_{\mathbf{k}} + \lambda^2 c^{(2)}_{\mathbf{k}} + ... && (2.52) \\ E(\mathbf{k}) &= E^{(0)}_{\mathbf{k}} + \lambda E^{(1)}_{\mathbf{k}} + \lambda^2 E^{(2)}_{\mathbf{k}} + ... && (2.53) \end{aligned}$$

We have already derived the zeroth order terms above

$$\begin{aligned} c^0_{n\mathbf{k}-\mathbf{G}} &= \delta_{\mathbf{G}-\mathbf{G}_n} && (2.54) \\ E^{(0)}_{n\mathbf{k}} &= E^0_{\mathbf{k}-\mathbf{G}_n} && (2.55) \end{aligned}$$

where we have now installed band indices to make clear that each band is associated with a particular reciprocal lattice vector.

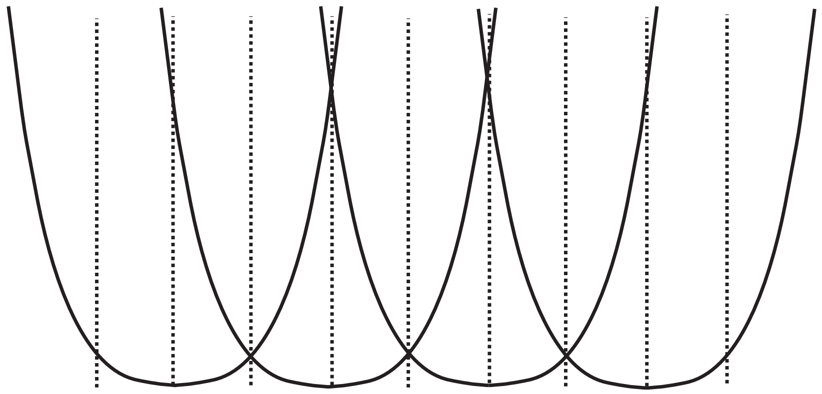


Figure 2.5: Free electron parabola in one dimension, with zone folding

The first order energy is just a trivial shift by the average potential $E^{(1)} = U_{\mathbf{0}}$, and the first order wavefunctions are

$$c^{(1)}_{n,\mathbf{k}-\mathbf{G}} = -\frac{U_{\mathbf{G}_n-\mathbf{G}}}{E^0_{\mathbf{k}-\mathbf{G}} - E^0_{\mathbf{k}-\mathbf{G}_n}} \qquad \text{for } \mathbf{G} \neq \mathbf{G}_n \ , \tag{2.56}$$

and the second-order energy correction is

$$E^{(2)}_{n,\mathbf{k}} = -\sum_{\mathbf{G}\neq\mathbf{G}_n} \frac{|U_{\mathbf{G}-\mathbf{G}_n}|^2}{E^0_{\mathbf{k}-\mathbf{G}} - E^0_{\mathbf{k}-\mathbf{G}_n}} \ . \tag{2.57}$$

One can, with patience, keep going to high order; in practice nowadays one just uses a large computer and diagonalises at will.

This equation tells us, as expected from our usual notions from perturbation theory, that non-degenerate bands repel each other, since every level that lies below the n^{th} band yields a positive energy shift, and every level above yields a negative shift. Notice, however, that there is a possibility that terms in the denominator of Eq. (2.56) and Eq. (2.57) vanish, which will lead to a singularity. This will occur at special values of momentum, whenever $\mathbf{k}$ lies on the zone boundary bisecting the reciprocal lattice vector $\mathbf{G} + \mathbf{G}_n$. Here, the perturbation theory fails for the obvious reason that we are coupling degenerate states, and non-degenerate perturbation theory does not hold. However small the perturbation, the mixing between two degenerate states is of order unity, and one must solve this degenerate problem exactly.

As an approximation, which will be valid near the degenerate points, let us ignore all coefficients except those corresponding to the degenerate plane waves. To prevent the notation getting out of hand, let us just focus on momenta near a single Brillouin zone boundary at $\mathbf{K}/2$, and consider $n = 0$. The (2x2) block out from Eq. (2.50) is just

$$\begin{pmatrix} E^0_{\mathbf{k}} - E & U_{\mathbf{K}} \\ U^*_{\mathbf{K}} & E^0_{\mathbf{k}-\mathbf{K}} - E \end{pmatrix} \begin{pmatrix} c_{\mathbf{k}} \\ c_{\mathbf{k}-\mathbf{K}} \end{pmatrix} = 0 \tag{2.58}$$

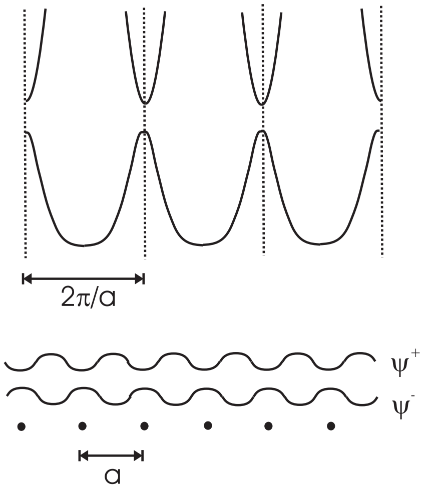


Figure 2.6: Energy bands in one dimension, and amplitudes of wavefunctions at the band edges

The solution of the determinantal equation leads to a quadratic equation:

$$E^{\pm}(\mathbf{k}) = \frac{1}{2}(E^0_{\mathbf{k}} + E^0_{\mathbf{k}-\mathbf{K}}) \pm \frac{1}{2}\sqrt{(E^0_{\mathbf{k}} - E^0_{\mathbf{k}-\mathbf{K}})^2 + 4|U_{\mathbf{K}}|^2} \tag{2.59}$$

It is instructive to look at the wavefunctions in this one-dimensional model. Exactly at $k = \frac{1}{2}K$, the energy levels are

$$E^{\pm}(\frac{1}{2}K) = E^0_{\frac{1}{2}K} \pm |U_K|, \tag{2.60}$$

and if we choose the potential to be attractive $U_K < 0$, the wavefunctions are

$$\begin{aligned} \psi^{-}(\frac{1}{2}K) &= \sqrt{2}\cos(\frac{1}{2}Kr) \ , \\ \psi^{+}(\frac{1}{2}K) &= \sqrt{2}\imath\sin(\frac{1}{2}Kr) \ . \end{aligned} \tag{2.61}$$

The wavefunctions are plotted, along with the potential, in Fig. 2.6.

The results are physically intuitive; because we chose the potential to be attractive, the lower energy state corresponds to wavefunctions concentrated on the potential minima (i.e. near the atoms) and the higher state pushes the charge density to the potential maxima in between the atoms. Of course, the potential in a real solid cannot be described by a single sine-wave, but the principle is clear.

One last comment is in order. Here we have just considered the two states on the zone boundary, which are strongly perturbed from plane waves. As one moves away from the boundary, the states become more nearly plane

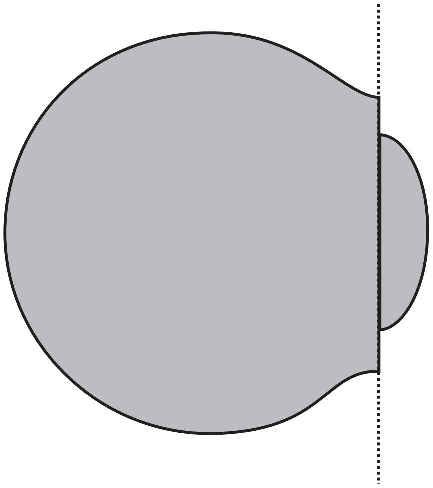


Figure 2.7: Fermi surface made by weak scattering from a single zone boundary

wave as one can see by inspection of Fig. 2.6, or from the energy eigenvalues. Clearly the k-states strongly affected are those for which

$$\frac{|E^0_{\mathbf{k}} - E^0_{\mathbf{k}-\mathbf{K}}|}{|U_{\mathbf{K}}|} = f \lesssim 1 \ . \tag{2.62}$$

States within this range will yield charge densities more-or-less like those of Eq. (2.61), whereas out of this range of momentum, they will be plane-wave-like and have a uniform charge density. So the amplitude of the periodic component of the total charge density will be of order f, which may be quite small. However, the states *exactly* at the zone boundary are always strongly perturbed.

Brillouin zones

Now we need to go back to discuss higher dimensions. We see from the above discussion that there will be a gap in the energy dispersion as one crosses a Brillouin zone boundary. But notice (see Eq. (2.59)) that if the momentum is changed parallel to the zone boundary, both $E^{\pm}(\mathbf{k})$ continue to disperse. To confirm this, let $\mathbf{k} = \mathbf{K}/2 + \mathbf{k}_\perp$, with $\mathbf{K} \cdot \mathbf{k}_\perp = 0$; then

$$E^{\pm}(\mathbf{k}_\perp) = \frac{\hbar^2}{2m}(\frac{K^2}{4} + k_\perp^2) \pm |U_{\mathbf{K}}| \tag{2.63}$$

Now consider a free electron Fermi sphere intersected by a single Brillouin zone boundary (see Fig. 2.7). Some reflection on Eq. (2.63) should convince you that if the scattering potential is weak, then the effect is to shift the Fermi surface outwards in below the gap, and inwards above the gap, as

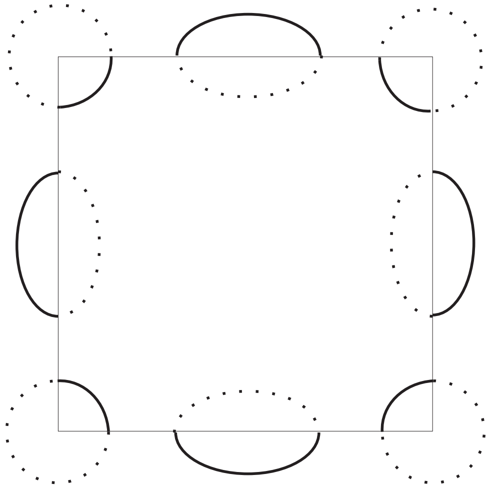


Figure 2.8: Fermi surface on a square lattice in the extended zone scheme. Solid lines show the Fermi surface in the first and second zones, and the dotted lines show the continuation into the extended zone scheme. In the reduced zone scheme one would show the two fermi surfaces within a single Brillouin zone

shown in the figure. Remember that the total area enclosed by the Fermi surface in k-space is conserved. If the potential is strong enough, the second Fermi sheet may shrink and vanish.

Here, the Fermi surface has been fractured into two pieces. If we had a simple square lattice, with a Fermi surface which crossed only the first Brillouin zone boundary, we would have the picture shown in Fig. 2.8. Here we have drawn a free-electron Fermi surface containing enough states to accommodate two electrons per unit cell. For small enough potential, the Fermi surface breaks up into two sheets, one in the first Brillouin zone, and one in the second. We can either draw the dispersion in the *extended zone scheme* (as here), or equivalently fold the dispersion back to make two bands in the *reduced zone scheme*. This picture makes it clear that the Fermi surface(s) are still *continuous*; here we have one Fermi surface which is *electron-like* and one which is *hole-like*. If we wished (and we shall), we could describe the material whose Fermi surfaces are given in this figure as a semimetal: a semiconductor, but with an overlap of the valence and conduction bands so that there are equal numbers of electrons and holes.

One can extend this type of construction to arbitrary complexity by including more reciprocal lattice vectors, each of which generates a new Brillouin zone boundary and further Brillouin zones. Even for nearly free electron metals, the Fermi surfaces which then are generated become quite complex[4]

---

[4] There are lots of pretty pictures in Ashcroft and Mermin (Chapter 9) and Marder

2.1.4 Tight binding description

In the last section, we calculated electronic bandstructure by viewing it as a gas of nearly plane wave excitations, with weak Bragg scattering by the crystal lattice. In this section, we shall discuss the opposite point of view; of a solid as a collection of weakly interacting neutral atoms.

Consider a periodic lattice of identical atoms, with an atomic Hamiltonian H_{atom} and a set of localised discrete levels with wavefunctions ϕ_n and energies E_n. Provided the overlap of the wavefunctions from one lattice site to the next is small, we might expect that a good approximation for the wavefunction in a solid will be to mix the degenerate levels from one atom and its neighbours.[5] Then we write an ansatz for a Bloch state

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{N^{\frac{1}{2}}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_n(\mathbf{r}-\mathbf{R}) \; . \tag{2.64}$$

where $\mathbf{R}$ are the reciprocal lattice vectors.

We can now evaluate the energy of the state in first order of perturbation theory:

$$\begin{aligned} E(\mathbf{k}) = < \mathbf{k}|H|\mathbf{k} > &= \frac{1}{N} \sum_{\mathbf{R}_j \mathbf{R}_m} e^{-i\mathbf{k}\cdot(\mathbf{R}_j-\mathbf{R}_m)} < \phi(\mathbf{r}-\mathbf{R}_m)|H|\phi(\mathbf{r}-\mathbf{R}_j) > \\ &= \sum_{\mathbf{R}_i} e^{-i\mathbf{k}\cdot\mathbf{R}_i} \int d\mathbf{r}\, \phi^*(\mathbf{r}-\mathbf{R}_i) H \phi(\mathbf{r}) \; . \end{aligned} \tag{2.65}$$

Because the wavefunctions are weakly overlapping, we will keep only the terms in Eq. (2.65) where the orbitals are on the same site, or on nearest neighbour sites connected by ρ. The onsite term gives an energy close to the atomic energy level, since $H \approx H_{atom}$ where the atomic wavefunction is concentrated, and we define

$$\epsilon_0 = \int d\mathbf{r}\, \phi^*(\mathbf{r}) H \phi(\mathbf{r}) \approx E_n \; , \tag{2.66}$$

and the term between neighbours - often called a *hopping integral* is

$$t = \int d\mathbf{r}\, \phi^*(\mathbf{r}-\rho) H \phi(\mathbf{r}) \; . \tag{2.67}$$

The band energy is then

$$E(\mathbf{k}) = \epsilon_0 + t \sum_{\rho} e^{-i\mathbf{k}\cdot\rho} \; . \tag{2.68}$$

(Chapter 8)

[5] For simplicity, we shall discuss the case when the atomic state is non-degenerate, i.e. an s- state, not p- or d-, which are more complicated, and drop the index n

As a simple example, take a cubic lattice with six nearest neighbours at a distance a. We then have

$$E(\mathbf{k}) = \epsilon_0 + 2t[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)] \ . \tag{2.69}$$

The bandwidth is $6t$, so that the more atomic-like, the narrower the band. At both the bottom and the top of the band, the dispersion is parabolic, i.e. free electron-(hole-) like, but with an effective mass $\hbar^2/2ta^2$; the narrower the band, the heavier the mass.

2.1.5 The pseudopotential

The NFE method and the tight-binding method are not accurate methods of electronic structure determination; nevertheless both of them exhibit the basic principles. They are commonly used to write down simple models for bands, with their parameters fit to more sophisticated calculations, or to experiment. It turns out that band gaps in semiconductors are usually fairly small, and the true dispersion can be modelled by scattering from a few Fourier components of the lattice potential. The reason is that the relevant scattering potential for valence band electrons is however MUCH smaller than the full atomic potential ze^2/r of an electron interacting with a nucleus of charge z. The effective potential for scattering of the valence electrons by the atomic cores is a weak *pseudopotential*.

When we consider the band structure of a typical solid, we are concerned only with the valence electrons, and not with those tightly bound in the core, which remain nearly atomic. If we solve the full Schrödinger equation with the real Coulomb potential, we expect to calculate not just the valence electronic states, but also the atomic like core states. A pseudopotential reproduces the valence states as the *lowest* eigenstates of the problem and neglects the core states.

A weak pseudopotential acting on a smooth pseudo-wavefunction gives nearly the same energy eigenvalues for the valence electrons as the full atomic potential does acting on real wavefunctions. Away from the atomic cores, the pseudopotential matches the true potential, and the pseudo-wavefunction approximates the true one.

A formal derivation of how this works can be given using the method of orthogonalised plane waves. The atomic states are well described by the Bloch functions $f_{n\mathbf{k}}$ of Eq. (2.64). Higher states, which extend well beyond the atoms will not necessarily be of this kind, but they must be *orthogonal* to the core levels. This suggests that we should use as a basis [6]

$$|\chi_\mathbf{k}>=|\mathbf{k}> - \sum_n \beta_n |f_{n\mathbf{k}}> , \tag{2.70}$$

---

[6] We use Dirac's notation of *bra* and *ket*, where $|\mathbf{k}>$ represents the plane wave state $\exp(i\mathbf{k}\cdot\mathbf{r})$, and $<\phi_1|T|\phi_2>$ represents the matrix element $\int d\mathbf{r}\, \phi_1^*(\mathbf{r})T(\mathbf{r})\phi_2(\mathbf{r})$ of the operator T.

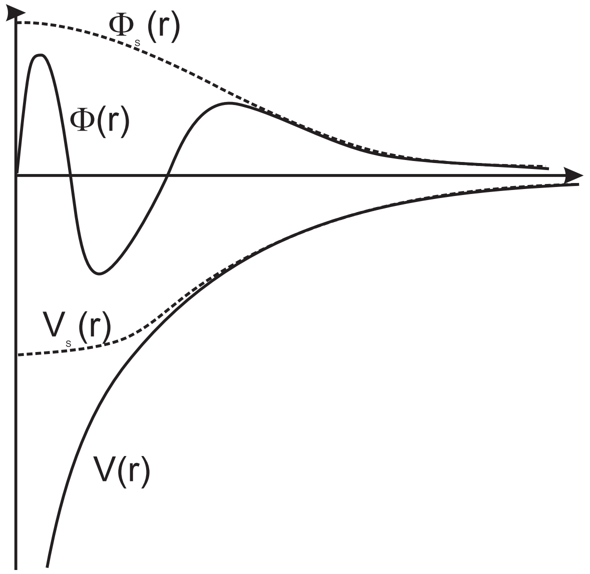


Figure 2.9: Pseudopotential: The true potential $V(r)$ has a wavefunction for the valence electrons that oscillates rapidly near the core. The pseudopotential $V_s(r)$ has a wavefunction $\Phi_s(r)$ that is smooth near the core, but approximates the true wavefunction far from the core region.

where $|\mathbf{k}>$ is a plane wave, and the coefficients $\beta_n(\mathbf{k})$ are chosen to make the states χ orthogonal to the core states $|f_{n\mathbf{k}}>$. The states in Eq. (2.70) with the coefficients determined in Question 2.2.5 are *orthogonalised plane waves* (OPW); away from the core, they are plane wave like, but in the vicinity of the core they oscillate rapidly so as to be orthogonal to the core levels.

We can now use the OPW's as basis states for the diagonalisation in the same way that we used plane waves in the NFE, viz

$$|\psi_k>=\sum_{\mathbf{G}}\alpha_{\mathbf{k}-\mathbf{G}}|\chi_{\mathbf{k}-\mathbf{G}}> \ . \tag{2.71}$$

This turns out to converge very rapidly, with very few coefficients, and only a few reciprocal lattice vectors are included in the sum. The following discussion explains why.

Suppose we have solved our problem exactly and determined the coefficients α. Now consider the sum of plane waves familiar from the plane-wave expansion, but using the same coefficients, i.e.

$$|\phi_k>=\sum_{\mathbf{G}}\alpha_{\mathbf{k}-\mathbf{G}}|\mathbf{k}-\mathbf{G}> , \tag{2.72}$$

and then [7] it is easily shown that

$$|\psi>=|\phi>-\sum_n<f_n|\phi>|f_n> \ . \tag{2.73}$$

Then substitute into the Schrodinger equation $H|\psi>=E|\psi>$, which gives us

$$H|\phi>+\sum_n(E-E_n)<f_n|\phi>|f_n>=E|\phi> \tag{2.74}$$

---

[7] Saving more notation by dropping the index k

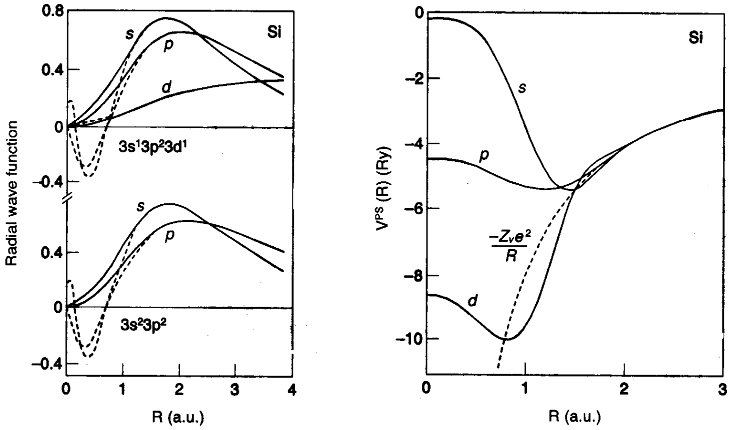


Figure 2.10: A comparison of the pseudowavefunction of Si with the corresponding all-electron wavefunctions for the configurations $3s^2 3p^2$ and $3s^2 3p^2 3d^1$, together with the corresponding psudopotential (for three different angular momentum states)[M.T.Yin and M.L.Cohen, *Phys.Rev.B* **26**, 5668 (1982)]

We may look upon this as a new Schrödinger equation with a *pseudopotential* defined by the operator

$$V_s|\phi> = U|\phi> + \sum_n (E - E_n) < f_n|\phi> \; |f_n> \tag{2.75}$$

which may be written as a non-local operator in space

$$(V_s - U)\phi(r) = \int V_R(\mathbf{r}, \mathbf{r}')\phi(\mathbf{r}')\, d\mathbf{r}' \;\; , \tag{2.76}$$

where

$$V_R(\mathbf{r}, \mathbf{r}') = \sum_n (E - E_n) f_n(\mathbf{r}) f_n^*(\mathbf{r}') \; . \tag{2.77}$$

The pseudopotential acts on the smooth *pseudo-wavefunctions* $|\phi>$, whereas the bare Hamiltonian acts on the highly oscillating wavefunctions $|\psi>$.

One can see in Eq. (2.75) that there is cancellation between the two terms. The bare potential is large and attractive, especially near the atomic core at $r \approx 0$; the second term V_R is positive, and this cancellation reduces the total value of V_s especially near the core. Away from the core, the pseudopotential approaches the bare potential.

The above are purely formal manipulations. The reason that the pseudopotential is useful is that the Fourier components of V_s are small, except for the first few reciprocal lattice vectors; and furthermore it is a good approximation to replace the pseudopotential by a local potential.[8] The pseudopotential is the formal justification for the NFE model.

---

[8] The latter restriction is not needed in general for modern electronic structure calculations

While this analytical formulation of the pseudopotential gives an important explanation of the success of the NFE method, it is almost never used in calculations. The usual formulation is to use a *model* pseudopotential for an atom which replaces each atomic potential by a weak potential which has the same scattering amplitude for the valence electrons. Question 2.13 shows how to describe a model bandstructure just in terms of reflection and transmission coefficients of a generalised atom. Nowadays, the most common is the *ab initio norm-conserving pseudopotential*, introduced by Hamann, Schlüter and Chiang [9]. We remarked above that the pseudopotential and the real potential are identical outside the core, and therefor the radial parts of the pseudo- wavefunction and the real wavefunction are proportional. The principle of the norm-conserving pseudopotential is to enforce *equality* of the wavefunctions outside the core – ensuring that the pseudo-charge density and true charge density are equal. An example of an atomic pseudopotential for Si is shown in Fig. 2.10.

2.2 Interactions

We have so far sidestepped entirely the effects of interactions between electrons, by working in the independent electron approximation with a one body potential $U(\mathbf{r})$ which somehow incorporates the interaction effects at the one particle level. Of course is is clear that the *full* Schrodinger equation Eq. (1.1) cannot be described by a set of one-body equations such as Eq. (2.12). However, we can ask the question as to what is the *best* single particle representation, and independently we can ask how good it is. This will be the subject of this section.

If we fix the position of the ions, the electronic Hamiltonian is

$$H_{elec} = \sum_{i=1}^{N} \left[-\frac{\hbar^2}{2m}\nabla_i^2 + U_{ion}(\mathbf{r}_i) \right] + \frac{1}{2}\sum_{i\neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{2.78}$$

where the potential due to the ions is

$$U_{ion}(\mathbf{r}) = \sum_I \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{2.79}$$

where Z_I is the nuclear charge and $\mathbf{R}_I$ the nuclear positions. We look for the solutions for the N-particle wavefunction $\Psi(\mathbf{r}_1, \sigma_1, ..., \mathbf{r}_N, \sigma_N)$ of the Schrödinger equation [10]

$$H_{elec}\Psi = E\Psi \; . \tag{2.80}$$

---

[9] D.R.Hamann, M.Schlüter, and C.Chiang, *Phys.Rev.Lett.* **43**, 1494 (1979)

[10] $\mathbf{r}_i$, σ_i are the space and spin coordinates of electron i

2.2.1 Preamble: a model two-electron system

The fundamental difficulty with treating interacting electronic systems is that we cannot expect to write a wavefunction that factorises into a product of single electron wavefunctions. Such a factorisation is disallowed by the required antisymmetry of wavefunctions for fermions, namely that

$$\Psi(\mathbf{r}_1,\sigma_1...,\mathbf{r}_i,\sigma_i,...,\mathbf{r}_j,\sigma_j,....,\mathbf{r}_N,\sigma_N) = -\Psi(\mathbf{r}_1,\sigma_1...,\mathbf{r}_j,\sigma_j,...,\mathbf{r}_i,\sigma_i,....,\mathbf{r}_N,\sigma_N) \quad . \tag{2.81}$$

Here $\mathbf{r}_i$,σ_i are the position and spin of particle i, and fermion wavefunctions change sign when the coordinates of any two electrons are interchanged.[11]

Almost everything that we shall do on the interacting system can be understood in simple terms for a model of an atom (or a molecule) with two single particle orbitals and two electrons. We shall assume that the (orthonormal) single particle states $\psi_{1,2}(\mathbf{r})$ are unchanged by the interaction. There is an important simplification that arises also because the Coulomb interaction between particles is independent of their spin state, which we shall denote by $|\uparrow>$, $|\downarrow>$. In that case, we already know that the eigenstates of the two particle problem should be labelled by the total spin S and its z-component S_z. There will be four possible spin states

$$\begin{array}{lll} & |\uparrow\uparrow> & \mathrm{S_z}=1 \\ \text{Triplet S} = 1 & (|\uparrow\downarrow> + |\downarrow\uparrow>)/\sqrt{2} & \mathrm{S_z}=0 \\ & |\downarrow\downarrow> & \mathrm{S_z}=-1 \end{array} \tag{2.82}$$

$$\text{Singlet S} = 0 \quad (|\uparrow\downarrow> - |\downarrow\uparrow>)/\sqrt{2} \quad \mathrm{S_z}=0 \tag{2.83}$$

The notation is that $|\uparrow\downarrow>=|\uparrow_1>|\downarrow_2>$, i.e. up spin for the electron labelled "1" and down spin for "2". The singlet state is *odd* under exchange of coordinates, and the triplet states are *even*. Because the total wavefunction must be odd, then the spatial wavefunctions that go along with these must be *odd* for the triplet states, and *even* for the singlet states. Since we decided at the outset that we are restricted to only two single particle states, we must have the following wavefunctions

$$\begin{aligned} \Psi_T(\mathbf{r}_1,\mathbf{r}_2) &= (\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_2(\mathbf{r}_1)\psi_1(\mathbf{r}_2))/\sqrt{2} \\ &= (|12> - |21>)/\sqrt{2} \qquad (2.84) \\ \Psi_S(\mathbf{r}_1,\mathbf{r}_2) &= (\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) + \psi_2(\mathbf{r}_1)\psi_1(\mathbf{r}_2))/\sqrt{2} \\ &= (|12> + |21>)/\sqrt{2} \qquad (2.85) \end{aligned}$$

where again we have used the notation that $|ij>=\psi_i(\mathbf{r}_1)\psi_j(\mathbf{r}_2)$, so particle "1" is in the spatial wavefunction labelled by the state "i", and particle "2" is

---

[11] Notice that the space and spin labels must both be interchanged.

in state "j". The subscripts S and T label singlet and triplet wavefunctions, respectively.

Notice that the antiymmetry of the triplet wavefunction means that the electrons keep further apart than they would if they were independent distinguishable particles; in the singlet (symmetric) state they are closer together. This means that we expect that the triplet state is lower in energy than the singlet state, given the Coulomb repulsion. This can be shown explicitly using by evaluating the expectation value of the Hamiltonian using the two wavefunctions, which is

$$< H >_{S,T}= E_1 + E_2 + < 12|V|12 > \pm < 21|V|12 > \quad , \tag{2.86}$$

where the $+/-$ signs are for singlet/triplet respectively. $E_{1,2}$ are the single particle energies - i.e. the expectation value of $T + U_{ion}$ - and the last two terms are matrix elements of the Coulomb interaction. The first of these is the direct, or Hartree energy

$$< 12|V|12 >= \int d\mathbf{r}d\mathbf{r}' |\psi_1(\mathbf{r}')|^2 \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} |\psi_2(\mathbf{r})|^2 \tag{2.87}$$

which is just the interaction energy between the charge densities of the two electronic single particle states. The second of these has no analogue classically, and is called the exchange energy

$$< 21|V|12 >= \int d\mathbf{r}d\mathbf{r}' \psi_2^*(\mathbf{r})\psi_1^*(\mathbf{r}') \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \psi_1(\mathbf{r})\psi_2(\mathbf{r}') \quad . \tag{2.88}$$

Despite the fact that the electron-electron interaction is independent of the spin of the electron, the requirement of antisymmetry of the wavefunction then produces a spin-dependent energy of the final state. As we shall see, this is the origin of magnetism in solids.

2.2.2 Hartree approximation

Returning to the many electron problem, we can now repeat the calculation more formally, but this time not restricting the single particle basis to predetermined states. The most natural first approximation to the difficult interaction term in Eq. (2.78) is to replace the interaction between the electrons by a set of interactions between a single electron and the charge density made up from all the other electrons, i.e. by a one-body potential for the i^{th} electron

$$U_{coul}(\mathbf{r}) = -e \int d\mathbf{r}' \, \frac{\rho(\mathbf{r}')}{\mathbf{r}-\mathbf{r}'} = e^2 \sum_{j \neq i} \int d\mathbf{r}' \, \frac{|\psi_j(\mathbf{r}')|^2}{\mathbf{r}-\mathbf{r}'} \ , \tag{2.89}$$

where the summation is over all the occupied states ψ_i. This clearly takes into account the averaged effect of the Coulomb repulsion due to all the other electron, and corresponds to the term Eq. (2.87) above.

It turns out that the Hartree approximation can also be derived as a variational theory. If we write as the ground state wavefunction as a product over orthonormal wavefunctions ψ_1, ψ_2, etc.

$$\Psi_{Hartree}(\mathbf{r}_1, \sigma_1, ..., \mathbf{r}_N, \sigma_N) = \psi_1(\mathbf{r}_1\sigma_1) \times ... \times \psi_N(\mathbf{r}_N, \sigma_N) \tag{2.90}$$

then the variational equations that minimise the total energy $< \Psi|H|\Psi >$ are

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U_{ion}(\mathbf{r}) + U_{coul}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \ . \tag{2.91}$$

These nonlinear Hartree equations can be solved, e.g. by iteration, to find the "best" set of variational wavefunctions ψ_i.

This Hartree approximation is an example of a *self-consistent field theory*, where the average configuration of all the other electrons affects the electron under consideration, but the particular *correlations* between any two (or more) electrons are ignored. It is not a simple exercise in practice to solve these equations, despite the crudity of the approximation.

2.2.3 Hartree-Fock

One of the primary deficiencies of the Hartree approximation is that the wavefunctions violate the Pauli principle: for fermions, the sign of the wavefunction must change when any two of the particles are exchanged, namely

$$\Psi_1(\mathbf{r}_1\sigma_1, ..., \mathbf{r}_i\sigma_i, ..., \mathbf{r}_j\sigma_j, ..., \mathbf{r}_N\sigma_N) = -\Psi_1(\mathbf{r}_1\sigma_1, ..., \mathbf{r}_j\sigma_j, ..., \mathbf{r}_i\sigma_i, ..., \mathbf{r}_N\sigma_N) \ . \tag{2.92}$$

This was precisely the physics that we were at pains to incorporate in the model two-electron problem above. The simplest wavefunction that satisfies this requirement is the *Slater determinant*

$$\Psi_{HF} = \begin{vmatrix} \psi_1(\mathbf{r}_1, \sigma_1) & \cdots & \psi_1(\mathbf{r}_N, \sigma_N) \\ \vdots & & \vdots \\ \psi_N(\mathbf{r}_1, \sigma_1) & \cdots & \psi_N(\mathbf{r}_N, \sigma_N) \end{vmatrix} \ . \tag{2.93}$$

If one evaluates the energy in the form

$$< H >_\Psi = \frac{< \Psi|H|\Psi >}{< \Psi|\Psi >}$$

with the determinantal wavefunction of Eq. (2.93) using an orthonormal set of orbitals ψ_i, one gets [12]:

$$< H >_\Psi = \sum_i < i|(T + U_{ion}|i > + \frac{1}{2}\sum_{ij}\left[< ij|\frac{e^2}{r_{ij}}|ij > - < ij|\frac{e^2}{r_{ij}}|ji > \delta_{\sigma_i\sigma_j}\right] \ . \tag{2.94}$$

---

[12] We shall use the notation $< i|f|j > = \int d\mathbf{r}\, \phi_i^*(\mathbf{r})f(\mathbf{r})\phi_j(\mathbf{r})$ for one body matrix elements, and $< ij|f|mn > = \int\int d\mathbf{r}\, d\mathbf{r}'\, \phi_i^*(\mathbf{r})\phi_j^*(\mathbf{r}')f(\mathbf{r}, \mathbf{r}')\phi_m(\mathbf{r})\phi_n(\mathbf{r}')$ for two-body matrix elements.

(This is much trickier than it looks). You should check that this result is consistent with the model two electron calculation above.

Then one can variationally minimise with respect to the ψ_i^* obtaining the Hartree-Fock equations

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + U_{ion}(\mathbf{r}) + U_{coul}(\mathbf{r})\right]\psi_i(\mathbf{r}) - \sum_j \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')\psi_j(\mathbf{r})\delta_{\sigma_i\sigma_j} = \epsilon_i\psi_i(\mathbf{r}) \tag{2.95}$$

After solving Eq. (2.95) to determine the wavefunctions and the energy eigenvalues the total energy can be written [13]

$$< H >_\Psi = \sum_i \epsilon_i - \frac{1}{2}\sum_{ij}\left[< ij|\frac{e^2}{r_{ij}}|ij > - < ij|\frac{e^2}{r_{ij}}|ji > \delta_{\sigma_i\sigma_j}\right] \ . \tag{2.96}$$

The equations are similar to the Hartree equations, but have an extra term, called the *exchange* term which is not only nonlinear in ψ but also *non-local*, and *spin-dependent.* This complexity makes them very difficult to use in practice.

The homogeneous electron gas

The one case where the Hartree-Fock equations can be solved straightforwardly is the not uninteresting case of jellium: the periodic lattice potential is replaced by a uniform positive background that neutralises the electronic charge. In this case the single particle states in the Slater determinant are just plane waves:

$$\psi_i(\mathbf{r}) = (V)^{-\frac{1}{2}} e^{i\mathbf{k}_i\cdot\mathbf{r}} \times \chi_{spin} \tag{2.97}$$

where we occupy each wavevector $|\mathbf{k}| < k_F$ twice (once for each spin component). Because we know by symmetry that the wavefunctions must have this form, we can evaluate the Hartree-Fock energy without having to solve for the wavefunctions, which is the hard bit in a problem with an atomic lattice.

It is useful at this point to introduce the electron gas density parameter r_s. Since the energy scale is set by the Coulomb potential, it is convenient to measure energies in units of the Rydberg:

$$1 Rydberg = \hbar^2/2ma_B^2 = e^2/2a_B \ , \tag{2.98}$$

with a_B the Bohr radius. Then we measure the density n in units of the Bohr radius by

$$4\pi(r_s a_B)^3/3 = 1/n \tag{2.99}$$

---

[13] Notice that this is not equal to the sum of single-particle energies, because otherwise the interaction terms would be counted twice

so r_s is a dimensionless parameter, which is the average spacing between electrons measured in units of the Bohr radius. You should be able to see that by dimensional analysis, r_s is the only relevant parameter.

The energy can be evaluated as follows. The direct Coulomb energy cancels exactly with the ionic energy: $U_{Coul} + U_{ion} = 0$. We are left with only the kinetic energy and the exchange energy. The kinetic energy term in Eq. (2.95) gives

$$-\frac{\hbar^2}{2m}\nabla^2 e^{i\mathbf{k}_i\cdot\mathbf{r}} = \frac{\hbar^2 k^2}{2m} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.100}$$

and the exchange term in Eq. (2.95) becomes

$$\begin{aligned} U_{exch} V^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}} &= -\sum_{|\mathbf{q}|<k_F} V^{-\frac{1}{2}} e^{i\mathbf{q}\cdot\mathbf{r}} \int d\mathbf{r}'\, V^{-\frac{1}{2}} e^{-i\mathbf{q}\cdot\mathbf{r}'} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} V^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}'} \\ &= -(V)^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}} \times \frac{1}{V} \sum_{q<k_F} \int d\mathbf{r}' e^{-i(\mathbf{k}-\mathbf{q})\cdot(\mathbf{r}-\mathbf{r}')} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \\ &= -(V)^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{r}} \times \frac{1}{V} \sum_{q<k_F} \frac{4\pi e^2}{|\mathbf{k}-\mathbf{q}|^2} \ . \end{aligned} \tag{2.101}$$

Eq. (2.101) makes use of the familiar result that the Fourier transform of the Coulomb potential $1/r$ is just $4\pi/q^2$.

The energy can be evaluated in closed form

$$\begin{aligned} \epsilon(\mathbf{k}) &= \frac{\hbar^2 k^2}{2m} - \int_{k'<k_F} \frac{d\mathbf{k}'}{(2\pi)^3} \frac{4\pi e^2}{|\mathbf{k}-\mathbf{k}'|^2} \\ &= \frac{\hbar^2 k^2}{2m} - \frac{2e^2 k_F}{\pi} F(k/k_F) \ , \end{aligned} \tag{2.102}$$

where

$$F(x) = \frac{1}{2} + \frac{1-x^2}{4x} \ln\left|\frac{1+x}{1-x}\right| \ . \tag{2.103}$$

The total energy can be performed by integration over $k < k_F$ It is [14]

$$\begin{aligned} E^{HF}_{tot} &= 2\sum_{k<k_F} \frac{\hbar^2 k^2}{2m} - \frac{2e^2 k_F}{\pi} \sum_{k<k_F} F(k/k_F) \\ &= N\left[\frac{3}{5}\frac{\hbar^2 k_F^2}{2m} - \frac{3}{4}\frac{e^2 k_F}{\pi}\right] \end{aligned} \tag{2.104}$$

This result is conventionally written in rydberg units, using the electron gas density parameter r_s (Eq. (2.99)) as

$$E^{HF}_{tot}/N = 2.21 r_s^{-2} - 0.916 r_s^{-1} \tag{2.105}$$

For a typical metal r_s is in the range of 2-4, and the second term is comparable in size to the first.

---

[14] Notice a factor of two for spin, and a factor of 1/2 (double-counting) in the second term

Note that the first term in Eq. (2.105) is from the kinetic energy, and the second from the interactions. The kinetic energy term is *dominant* at smaller r_s, which means higher density - despite the electrons coming closer to each other. This is of course precisely the opposite result that one would have got for a classical gas (where the "perfect gas" limit is the *dilute* limit. The difference is because electrons are fermions, and as one goes to higher density, the Fermi energy E_F increases more rapidly than the interaction energy. *High density metals are free-electron like.*

Eq. (2.105) looks like the first few terms in a series expansion starting from the high density limit. The series continues[15]

$$E_{tot}/N = 2.21 r_s^{-2} - 0.916 r_s^{-1} + 0.0622 \ln(r_s) - 0.096 + O(r_s) + ... \quad (2.106)$$

The difference in energy between the exact energy and the Hartree-Fock energy is often termed the *correlation energy.*

Wigner crystal

We discussed above that interactions become relatively more important as the density decreases, because the kinetic energy penalty for keeping electrons apart is reduced. But we are still describing the electrons as a fluid, albeit a quantum fluid; our experience tells us that if repulsive interactions get large enough, the ground state of the system should not be a fluid but a solid. Quite some time ago, Wigner[16] argued that the electron gas must crystallise at low enough density. The argument is quite simple, but quite beyond the reach of the independent electron approximation.

In an electron solid, each electron is assumed to be localised to a site on a lattice, not spread out through the crystal as in the metallic state we have been discussing hitherto. If the electrons are well localised and only weakly overlap with each other, one can estimate the energy quite simply.

The Coulomb energy is just the energy of interacting charged particles - the Hartree term. If you include the Jellium background, the interaction of with the N-1 other electrons is approximately cancelled by their counterbalancing background charge. What remains is the Coulomb energy of a single electron that has a uniformly charged sphere of jellium around it. This is of order

$$-\frac{e^2}{2a_B r_s} \quad (2.107)$$

(One can improve the estimate by doing the calculation summing over all the charges; this multiplies the result by a factor of α, called a Madelung constant. Generally $\alpha \approx 1.7$ depending on the crystal lattice structure).

---

[15] Gell-Mann and Brueckner, *Phys. Rev.* **106**, 364 (1957).

[16] E.P.Wigner, *Phys. Rev.* **46**, 1002 (1934)

The kinetic energy cost of localising the electron is just the zero point energy of localising the electron inside the sphere which should be of order

$$\frac{\hbar^2}{2ma_B^2 r_s^2} \tag{2.108}$$

Thus we estimate

$$E_{tot}^{WC}/N \approx -\frac{1}{2}\alpha r_s^{-1} + r_s^{-2}\ Ryd. \tag{2.109}$$

For large enough r_s, the coulomb energy must always win – because asymptotically the dominant energy term is negative, and scales as r_s^{-1}.[17] Comparing Eq. (2.109) with Eq. (2.106) suggests a transition near $r_s \approx 5$, whereas the best calculations place it closer to 100.[18]

2.2.4 Density functional theory

We have spent a lot of time discussing qualitatively useful but quantitatively inaccurate methods for treating interactions. We close this chapter by discussing a *practical* method for including the effects of interactions, that is based on a remarkable theorem.

The Hohenberg-Kohn theorem

The density functional theorem states that the total ground state energy of the interacting electron system (including the interaction with the ions) is a functional of the electronic density $n(\mathbf{r})$, and that the ground state density and energy can be obtained by minimising this functional with respect to the density. This is a surprising result, since it somehow avoids discussing wavefunctions (and quantum mechanics directly) at all. The hitch is that the functional is unknown in general; the practical advance is that very good approximations to the functional are known.

How this comes about, we shall now sketch. There are in fact two theorems; or a theorem and a corollary, to be pedantic.

We first rearrange the notation of Eq. (2.78) as

$$H = H_{int} + V_{ext} \tag{2.110}$$

where

$$H_{int} = T + V_{ee} \tag{2.111}$$

---

[17] Notice that Eq. (2.109) is an expansion in powers of $(1/r_s)$ about the *low-* density limit ($r_s \to \infty$), whereas Eq. (2.106) is an expansion in powers of r_s about the *high*-density limit ($r_s \to 0$).

[18] The most reliable estimates come from quantum Monte Carlo calculations, see D.Ceperley and B.J.Alder, *Phys. Rev. Lett.* **45**, 566 (1980).

the sum of the kinetic energy term and the electron-electron interaction term, and

$$V_{ext} = U_{ion} = \sum_i v_{ext}(\mathbf{r}_i) \tag{2.112}$$

The interaction with the ions is viewed as an external potential acting on the electronic system; $\mathbf{r}_i$, $i = (1, 2, ..., N)$ are the coordinates of the electrons.

The electronic density corresponding to the ground state many-body wavefunction is

$$\begin{aligned} n(\mathbf{r}) &= < \Psi_0(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)| \sum_i \delta(\mathbf{r} - \mathbf{r}_i) |\Psi_0(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) > \\ &= N \int d\mathbf{r}_1 ... d\mathbf{r}_N \, |\Psi_0(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)|^2 \delta(\mathbf{r} - \mathbf{r}_1) \end{aligned} \tag{2.113}$$

Note that

$$< \Psi_0 | V_{ext} | \Psi_0 > = \int d\mathbf{r} \, n(\mathbf{r}) v_{ext}(\mathbf{r}) \ , \tag{2.114}$$

using Eq. (2.113); we need only the density, and not the wavefunctions, to evaluate this term in the energy.

Notice that the interaction energy between the electrons and the Let us imagine that we hold H_{int} fixed, as well as the total number N of electrons, and vary the "external" potential V_{ext}. The first Hohenberg-Kohn theorem states that *there is a one-to-one correspondence between the ground-state density of a N-electron system and the external potential acting on it.* This theorem implies that the electron density contains *in principle* all the information contained in the ground state wavefunction.

Proof:

To prove the theorem, assume it is false. Assume there are two potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$ that result in the same ground state density $n(\mathbf{r})$. The two resulting Hamiltonians $\hat{H}_1$ and $\hat{H}_2$ will then have ground state wavefunctions Ψ_1 and Ψ_2, and energies E_1 and E_2. The variational principle tells us that[19]

$$E_1 = < \Psi_1 | \hat{H}_1 | \Psi_1 > < < \Psi_2 | \hat{H}_1 | \Psi_2 > \ , \tag{2.115}$$

and so

$$\begin{aligned} E_1 &< < \Psi_2 | \hat{H}_2 | \Psi_2 > + < \Psi_2 | (\hat{H}_1 - \hat{H}_2) | \Psi_2 > \\ E_1 &< E_2 + \int d\mathbf{r} \, n(\mathbf{r}) [v_1(\mathbf{r}) - v_2(\mathbf{r})] \ . \end{aligned} \tag{2.116}$$

However, the argument can easily be run in the reverse order, switching the indices 1 and 2, viz.

$$E_2 < E_1 + \int d\mathbf{r} \, n(\mathbf{r}) [v_2(\mathbf{r}) - v_1(\mathbf{r})] \ . \tag{2.117}$$

---

[19] We assume the ground states of the two Hamiltonians are not degenerate, which is a technicality that can easily be fixed

The equations Eq. (2.116) and Eq. (2.117) are in contradiction unless $v_1 = v_2$ (and then of course $E_1 = E_2$).

The result of the theorem is to say that the external potential is a *functional* of the density: we express this relationship with square brackets $v_{ext}[n]$.

In order to use the theorem, we must make a few obvious remarks. The ground state wavefunction is a functional of the external potential v_{ext}, and by extension this follows for the energy $E[v_{ext}]$, the kinetic energy $T[v_{ext}]$ and the interaction energy $V_{ee}[v_{ext}]$. But since we have seen that v_{ext} is a functional of the density $n(\mathbf{r})$, all of these are functionals of the density, as is the ground state wavefunction itself.

This brings us to the second Hohenberg-Kohn theorem, often called the density functional theorem, which is just the formulation of a variational principle on the functional

$$\begin{aligned} E^{HK}[n(\mathbf{r}); v_{ext}(\mathbf{r})] &= \; < \Psi_0[n]|T + V_{ee} + V_{ext}|\Psi_0[n] > \\ &= \; T[n(\mathbf{r})] + V_{ee}[n(\mathbf{r})] + \int d\mathbf{r}\, n(\mathbf{r}) v_{ext}(\mathbf{r}) \end{aligned} \tag{2.118}$$

where v_{ext} is now taken to be *fixed*, and $n(\mathbf{r})$ is allowed to vary, with $|\Psi_0[n] >$ the ground state wavefunction of a system with density $n(\mathbf{r})$. We know that a variational principle exists with regard to the wavefunctions $|\Psi >$, but because of the functional dependence of $|\Psi >$ on n, a variational minimisation of of E^{HK} with respect to $n(\mathbf{r})$ yields the *exact* ground state density, and the *exact* ground state energy. Notice furthermore (and this is the useful bit) that the function $F[n] = T[n] + V_{ee}[n]$ is *universal* and does not depend on v_{ext}(which contains all the information about the actual solid) – the problem once solved is solved for ever. Unfortunately, the functional $F[n]$ is not known of course – but in many cases it can be approximated very well.

The Kohn-Sham equations

To carry out the variational minimisation it is convenient to decompose the density into a sum of *fictitious* independent orbital contributions

$$n(\mathbf{r}) = \sum_i^N \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}) \; . \tag{2.119}$$

where the ϕ_i are orthonormal [20]. These wavefunctions are *not* the wavefunctions for the interacting system – this is just a useful trick.

---

[20] We are allowed to do this by use of the Hohenberg-Kohn theorem. Imagine a noninteracting system, whose ground state density $n_0(\mathbf{r})$ equals $n(\mathbf{r})$. The HK theorem guarantees the uniqueness of a potential v_{ext}^0 that produces the ground state density n_0. But because the system is non-interacting, the ground state is described by a Slater determinant of the wavefunctions ϕ_i. So the decomposition of Eq. (2.119) exists.

We then write the variational function (Eq. (2.118)) as

$$E^{HK}[n] = T_0[n(\mathbf{r})] + V_H[n(\mathbf{r})] + \int d\mathbf{r}\, n(\mathbf{r}) v_{ext}(\mathbf{r}) + E_{xc}[n] \; , \tag{2.120}$$

where for convenience we have removed the Hartree energy

$$V_H[n] = \frac{1}{2} \int\int d\mathbf{r}\, d\mathbf{r}'\, n(\mathbf{r}) n(\mathbf{r}') \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} = \frac{1}{2} \sum_{ij} < ij|V_{ee}|ij > \; , \tag{2.121}$$

and the kinetic energy is now that of a system of non-interacting electrons

$$T_0[n] = \sum_i < i| - \frac{\hbar^2 \nabla^2}{2m} |i> \; . \tag{2.122}$$

The remaining energy is lumped into the *exchange-correlation* functional

$$E_{xc}[n] = T[n] - T_0[n] + V_{ee}[n] - V_H[n] \tag{2.123}$$

which is, of course, unknown.

Now we use the standard variational procedure on the N orbitals $\{\phi_i\}$, obtaining the *Kohn-Sham* equations

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + U_{ion}(\mathbf{r}) + U_{coul}(\mathbf{r}) + U_{xc}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \; , \tag{2.124}$$

where U_{ion} and U_{coul} are just the nuclear and Hartree potentials, as before and the exchange-correlation potential is formally defined as

$$U_{xc} = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \; . \tag{2.125}$$

After these equations have been solved, the exact ground state energy is given by

$$E_0 = \sum_i \epsilon_i - \frac{1}{2} \sum_{ij} < ij| \frac{e^2}{r_{12}} |ij> + E_{xc}[n] - \int d\mathbf{r}\, U_{xc}(\mathbf{r}) n(\mathbf{r}) \; . \tag{2.126}$$

The energies ϵ_i are purely formal Lagrange multipliers: any identification with one-particle energies is purely heuristic. Of course, in situations where the exchange-correlation energy is small (so called weakly correlated materials, which includes many common metals and semiconductors) they are a good approximation to single-particle terms.

The formal theory is of no practical use without a knowledge of the exchange-correlation potential U_{xc}. However, here we strike good fortune. It turns out that in many cases a very good approximation is to replace it by that calculated for a *uniform* electron gas of the same local density:

$$E_{xc}^{LDA}[n(\mathbf{r})] = \int d\mathbf{r} \epsilon_{xc}(n(\mathbf{r}))\, n(\mathbf{r}) \; . \tag{2.127}$$

where $\epsilon_{xc}(n(\mathbf{r}))$ is the exchange-correlation energy per electron. This is the *local density approximation* (LDA) which is widely used for accurate *ab initio* calculations. It could be taken from the series expansion in Eq. (2.106), for example, although there are better interpolations schemes over a wide density range, and some that incorporate gradient corrections in the density.

2.2.5 Screening and Thomas-Fermi theory

One of the most important manifestations of electron-electron interactions in metals is that of screening. If we insert a positive test charge into a metal, it attracts a cloud of electrons around it, so that at large distances away from the test charge the potential is perfectly screened. We shall be looking at screening later in the context of response functions, but we can look at a simple, and very useful, model here, that goes by the name of Thomas-Fermi theory in metals. It will be derived in two ways: firstly the traditional physical route, and secondly (but completely equivalently) couched in the language of density functional theory. To simplify matters, we will ignore entirely the exchange correlation potential entirely, so the energy consists of kinetic, Hartree, and external potential terms.

General remarks on screening

The charge density we introduce into the solid we will call $\rho_{ext}(\mathbf{r}) = |e|n_{ext}(\mathbf{r})$. In vacuum, or if for a moment we stop the electrons in the solid from responding to this charge, it would produce a potential v_{ext} that satisifies Poisson's equation

$$\nabla^2 v_{ext}(\mathbf{r}) = 4\pi e^2 n_{ext}(\mathbf{r}) \quad . \tag{2.128}$$

Once the electrons in the solid relax to accommodate the new potential, the total charge density will consist of the external charge and the induced electron charge density $n_{ind}(\mathbf{r})$, viz

$$\rho_{tot}(\mathbf{r}) = \rho_{ext}(\mathbf{r}) + \rho_{ind}(\mathbf{r}) = e(n_{ext}(\mathbf{r}) - n_{\mathbf{r}}(\mathbf{r})) \quad , \tag{2.129}$$

which generates the actual potential v_{tot} seen by the electrons,

$$\nabla^2 v_{tot}(\mathbf{r}) = 4\pi e^2 (n_{ext}(\mathbf{r}) - n_{ind}(\mathbf{r})) \quad . \tag{2.130}$$

Because the electrons move toward the positive charge, the net effect will be that the total potential seen by an individual electron in the Schrodinger equation is less than the external potential.

In general, this phenomenon is incorporated into electromagnetic theory through the dielectric function ϵ, and usually one assumes that the total potential and the induced potential are linearly related (linearity being an assumption, for the moment). The dielectric function relates the electric displacement D to the electric field E, in the form

$$\mathbf{D}(\mathbf{r}) = \int d\mathbf{r}'\ \epsilon(\mathbf{r},\mathbf{r}')\mathbf{E}(\mathbf{r}') \quad . \tag{2.131}$$

Because $\mathbf{D}$ is generated by "free" charges (i.e. ρ_{ext}) and $\mathbf{E}$ by the total charge (i.e. ρ_{tot}) Eq. (2.131) can be rewritten as a relationship between the potentials generated by those charge distributions:

$$v_{ext}(\mathbf{r}) = \int d\mathbf{r}'\ \epsilon(\mathbf{r},\mathbf{r}')v_{tot}(\mathbf{r}') \quad . \tag{2.132}$$

In a spatially uniform electron gas, ϵ can depend only on the separation between coordinates, i.e. $\epsilon(\mathbf{r},\mathbf{r}') = \epsilon(\mathbf{r}-\mathbf{r}')$, so that Eq. (2.132) becomes a convolution in real space – better written of course in Fourier (momentum) space as[21]

$$v_{ext}(\mathbf{q}) = \epsilon(\mathbf{q})v_{tot}(\mathbf{q}) \quad . \tag{2.133}$$

The job of a microscopic model is to calculate the induced charge density (using the Schrodinger equation) produced by the external potential, i.e. $n_{ind}([v_{ext}(\mathbf{r})])$; again assuming it is linear (expected to be the case if v_{ext} is small enough, this will be a calculation of what is usually called the *linear susceptibility* χ, defined by

$$n_{ind}(\mathbf{q}) = \chi(\mathbf{q})v_{ext}(\mathbf{q}) \quad . \tag{2.134}$$

Looking back to the definitions of the potential in terms of the density, the relationship to the dielectric function is

$$\frac{1}{\epsilon(\mathbf{q})} = 1 - \frac{4\pi e^2}{q^2}\chi(\mathbf{q}) \quad . \tag{2.135}$$

Thomas-Fermi approximation

The Thomas-Fermi theory of screening starts with the Hartree approximation to the Schrödinger equation. We shall just treat the case of “jellium”, where the ionic potential is spread out uniformly to neutralise the electron liquid. So the only potential in the problem is the *total* potential (external plus induced) produced by the added charge and by the non-uniform screening cloud (see Fig. 2.11)

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + v_{tot}(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad . \tag{2.136}$$

---

[21] Here we use the same symbol for a function and its Fourier transform, distinguishing the two by their arguments

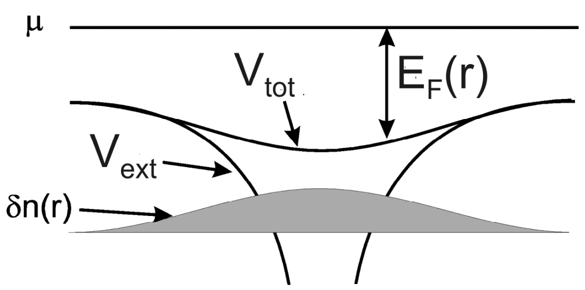


Figure 2.11: Thomas-Fermi approximation

The second approximation is to assume that the induced potential is slowly varying enough that the energy eigenvalues of Eq. (2.136) are still indexed by momentum, but just shifted locally by the potential:

$$E(\mathbf{k},\mathbf{r}) = \frac{\hbar^2 k^2}{2m} + v_{tot}(\mathbf{r}) \quad . \tag{2.137}$$

This makes only sense in terms of wavepackets, but provided the potential varies slowly enough on the scale of the Fermi wavelength $2\pi/k_F$, this approximation is reasonable. We know that after the system comes to equilibrium the chemical potential must be a constant, throughout the system. Keeping the electron states filled up to a constant energy μ requires that we adjust the local k_F such that

$$\mu = \frac{\hbar^2 k_F(\mathbf{r})^2}{2m} + v_{tot}(\mathbf{r}) = E_F(\mathbf{r}) + v_{tot}(\mathbf{r}) \quad , \tag{2.138}$$

where we have adopted the usual convention of measuring E_F from the bottom of the parabolic band[22].

Now we can substitute into Eq. (2.138) for the two terms on the RHS. We assume that k_F just depends on the local density

$$k_F(\mathbf{r}) = (3\pi^2 n(\mathbf{r}))^{1/3} \tag{2.139}$$

and we have from the discussion above that

$$v_{tot}(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_{ind}(\mathbf{r}) = v_{ext}(\mathbf{r}) + \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} n(\mathbf{r}') \tag{2.140}$$

Putting this all together, we get the Thomas-Fermi equation

$$\frac{\hbar^2}{2m}(3\pi^2)^{2/3} n^{2/3}(\mathbf{r}) + \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} n(\mathbf{r}') = \mu - v_{ext}(\mathbf{r}) \, . \tag{2.141}$$

---

[22] One is often sloppy about using E_F and μ interchangeably; here is a place to take care

This is still a nonlinear equation, but we shall be interested in the case when the added potential v_{ext} is small, so the density cannot differ very much from the density n_o of the system without the potential. Let us then write

$$\delta n(\mathbf{r}) = n(\mathbf{r}) - n_o \ , \tag{2.142}$$

where δn is the induced charge density. Linearising Eq. (2.141) we get

$$\frac{\hbar^2}{3m}(3\pi^2)^{2/3} n_o^{-1/3} \delta n(\mathbf{r}) + \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \delta n(\mathbf{r}') = -v_{ext}(\mathbf{r}) \tag{2.143}$$

This is a linear integral equation that is best solved by Fourier transforms. The solution can be written as

$$\delta n(\mathbf{q}) = -\frac{v_{ext}(\mathbf{q})}{\frac{4\pi e^2}{q^2}\left[1 + \frac{q^2}{q_{TF}^2}\right]} \ , \tag{2.144}$$

where the Thomas-Fermi wavevector is

$$q_{TF}^2 = \frac{4}{\pi}\frac{me^2}{\hbar^2} k_F = \frac{4}{\pi}\frac{k_F}{a_B} = \left(\frac{2.95}{r_s^{\frac{1}{2}}} \mathring{A}^{-1}\right)^2 . \tag{2.145}$$

Eq. (2.144) gives us the definition of the density response function χ, and hence we have calculated the static dielectric function in the Thomas-Fermi approximation

$$\epsilon^{TF}(q) = 1 + \frac{q_{TF}^2}{q^2} \quad , \tag{2.146}$$

or equivalently

$$v_{tot}(q) = v_{ext}(q) \frac{q^2}{q^2 + q_{TF}^2} \ . \tag{2.147}$$

In particular, if $v_{ext} = Q/r$ is Coulombic (long range), $V(r) = (Q/r)e^{-q_{TF} r}$ is a short-range Yukawa, or screened potential[23].

In a typical metal, r_s is in the range $2-6$, and so potentials are screened over a distance comparable to the interparticle spacing; the electron gas is highly effective in shielding external charges.

Thomas-Fermi is a density functional theory

The theory in the last section is all written as variational equations in terms of the density $n(\mathbf{r})$, which hints that the Thomas-Fermi approximation is in fact no more than a particular simple approximation to the density functional theory of Sec. 2.2.4. Here is the result derived in that fashion.

---

[23] This form is originally due to P.Debye and E.Hückel, *Zeitschrift für Physik* **24**, 185, (1923)and was derived for the theory of electrolytes; it appears also in meson theory under the name of the Yukawa potential; the physics in all cases is identical

For the kinetic energy, we just use the free-electron result for a uniform electron gas (Eq. (2.104)) for the kinetic energy density (per unit volume)

$$T = \frac{3}{5}\frac{\hbar^2 k_F^2}{2m} n = \frac{3}{5}\frac{\hbar^2}{2m}(3\pi^2)^{2/3} n^{5/3} \tag{2.148}$$

and then write an approximate funcional by assuming this is good for a slowly varying density, i.e.

$$T[n] = \int d\mathbf{r}\, \frac{3}{5}\frac{\hbar^2}{2m}(3\pi^2)^{2/3} n^{5/3}(\mathbf{r}) \ , \tag{2.149}$$

and the Hartree term is as before

$$V_H[n] = \frac{1}{2}\int\int d\mathbf{r}\, d\mathbf{r}'\, n(\mathbf{r})n(\mathbf{r}')\frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \ , \tag{2.150}$$

whereas the external potential is the interaction of the density with our true external potential that we have added to the system

$$V_{ext}[n] = \int d\mathbf{r}\, n(\mathbf{r}) v_{ext}(\mathbf{r}) \ . \tag{2.151}$$

We now minimise the functional $E^{TF} = T + V_H + V_{ext}$ with respect to n, subject to the constrain that the total number of particles is conserved. Formally, one introduces a Lagrange multiplier μ (otherwise known as the chemical potential), and thus get

$$\frac{\delta E[n]}{\delta n} = \mu \tag{2.152}$$

which generates the Thomas-Fermi equation of Eq. (2.141)[24]

$$\frac{\hbar^2}{2m}(3\pi^2)^{2/3} n^{2/3}(\mathbf{r}) + \int d\mathbf{r}'\, \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} n(\mathbf{r}') = \mu - v_{ext}(\mathbf{r}) \ . \tag{2.153}$$

---

[24] It is straightforward to also include the exchange term from Eq. (2.104), and the resultant equation is then called the Thomas-Fermi-Dirac equation

Questions on chapter 2

Qu.2.1 BCC and FCC lattices

Show that the reciprocal lattice of a body centred cubic lattice (BCC) of spacing a is a face centred cubic (FCC) lattice of spacing $4\pi/a$; and that the reciprocal lattice of a FCC lattice of spacing a is a BCC lattice of spacing $4\pi/a$.

Qu.2.2 Reciprocal lattice

Show that the volume of the primitive unit cell of the reciprocal lattice is $(2\pi)^3/\Omega_{cell}$, where Ω_{cell} is the volume of the primitive unit cell of the crystal.

Qu.2.3 * Another proof of Bloch's theorem

A more elegant way to prove Bloch's theorem is to note that the translation operator can be written

$$T_{\mathbf{R}} = e^{-i\hat{\mathbf{P}}\cdot\mathbf{R}/\hbar} ,$$

where $\hat{\mathbf{P}}$ is the momentum operator. (If you don't know how to do this, make the replacement $\hat{\mathbf{P}} = -i\hbar\nabla$, and check that the operator generates the infinite Taylor series expansion of $f(\mathbf{r}+\mathbf{R})$.) By multiplying by the ket $< \mathbf{k}|$ (an eigenfunction of momentum), show that either $< \mathbf{k}|\psi >= 0$, or $c(\mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{R}}$.

Qu.2.4 Density of states for free electrons

Calculate the density of states in energy for electrons with a quadratic dispersion $E = \hbar^2k^2/2m$ in one, two, and three dimensions.

[Answer:$(2m/\pi\hbar^2) \times (\hbar^2/2mE)^{\frac{1}{2}}$, (d=1); $(m/\pi\hbar^2)$, d=2; $(m/\pi^2\hbar^2) \times (2mE/\hbar^2)^{\frac{1}{2}}$, d=3 .]

Show how in three dimensions, this can be re-written as

$$(3/2)(n/E_F)(E/E_F)^{\frac{1}{2}}$$

with $n = N/V$.

Qu.2.5 One-dimensional band

Consider a one-dimensional system which is filled up to the first Brillouin zone boundary at $k = \pi/a$, and assume that there is a small gap produced by a single Fourier component of the lattice potential $U = U_{K=2\pi/a}$ (small meaning that $U/E^0_{\frac{1}{2}K} \ll 1$). Consider momenta close to the zone boundary, show that a good approximation for the energy dispersion of the bands is

$$E = E_0\left(1 \pm \sqrt{\frac{U^2}{E_0^2} + 4x^2}\right)$$

where $E_0 = E^0_{\frac{1}{2}K}$ and $k = (\pi/a)(1+x)$, with $x \ll 1$.

Evaluate the change in electronic energy

$$E_{elec} = \sum_{k\ occupied} [E(k;U_K) - E(k;U_K = 0)]$$

to leading order in U/E_0. (Hint: the result is non-analytic - **don't** try a Taylor series in U.)
We shall be using this result later to understand the Peierls instability of a one-dimensional system.

Qu.2.6 Some one-dimensional chemistry

Consider a diatomic lattice of two atoms labelled A and B in a lattice with period a, at the positions $\pm a/4(1-\delta)$ in a one-dimensional array with overall period a.

Using the NFE approximation valid for momenta near the zone boundary $k \to \pi/a$, show that the solution of Eq. (2.58) leads to:
(a) a gap on the zone boundary is $2|U_{2\pi/a}|$, and
(b) wavefunctions that satisfy $c^{\pm}_k/c^{\pm}_{k-\pi/a} = \pm U/|U|$ as $\mathbf{k} \to \pi/a$.

Hence show that the probability density for the electronic states at $k = \pi/a$ take the form

$$\begin{aligned} |\psi^+(r)|^2 &\propto \cos^2(\frac{\pi x}{a} + \frac{\phi}{2}) \\ |\psi^-(r)|^2 &\propto \sin^2(\frac{\pi x}{a} + \frac{\phi}{2}) \ . \end{aligned} \tag{2.154}$$

Show that the potential can be written

$$U_{2\pi/a} = \sin(\frac{\pi\delta}{2})(U^A_{2\pi/a} + U^B_{2\pi/a}) - i\cos(\frac{\pi\delta}{2})(U^A_{2\pi/a} - U^B_{2\pi/a}) \ , \tag{2.155}$$

where

$$U^{A,B}_{2\pi/a} = \frac{N}{V}\int dr\, e^{-2\pi i r/a} U^{A,B}(r) \ , \tag{2.156}$$

The system contains an average of one electron per atom, or equivalently two electrons per unit cell. Discuss the values of the energy gaps and plot the charge densities corresponding to the highest filled electron state and the lowest empty electron state in the two cases; (a) $\delta = 0$, $U_A \neq U_B$; (b) identical atoms, $U_A = U_B$, and $\delta \neq 0$.

Explain how this provides a simple model of either an *ionic* or *covalent* solid.

Qu.2.7 Tight binding for BCC and FCC lattices

Show that the tightbinding bandstructure for a body centred cubic lattice (include only the hopping to the eight nearest neighbours) is

$$E(\mathbf{k}) = \epsilon_0 + 8t\cos(\frac{1}{2}k_x a)\cos(\frac{1}{2}k_y a)\cos(\frac{1}{2}k_z a) \ , \tag{2.157}$$

and for the face centred cubic lattice (twelve nearest neighbours)

$$E(\mathbf{k}) = \epsilon_0 + 4t[\cos(\frac{1}{2}k_x a)\cos(\frac{1}{2}k_y a) + \cos(\frac{1}{2}k_y a)\cos(\frac{1}{2}k_z a) + \cos(\frac{1}{2}k_z a)\cos(\frac{1}{2}k_x a)] \ . \tag{2.158}$$

Qu.2.8 * Pseudopotential

Show that $<\chi|f_n>=0$ if we choose $\beta_n =< f_n|\mathbf{k}>$.

The pseudopotential is not unique. Show that the valence eigenvalues of a Hamiltonian $H + V_R$ are the same for any operator of the form

$$V_R\phi = \sum_n < F_n|\phi > f_n \ ,$$

where the F_n are **arbitrary** functions.

Qu.2.9 Hartree-Fock theory for the two level atom

Show that the Hartree-Fock total energy Eq. (2.94) applied to the two-level atom model of Sec. 2.2.1 gives exactly the direct and exchange energy calculated in Eq. (2.86) .

Qu.2.10 * Hartree Fock equations

This problem is just to derive the Hartree-Fock equations shown above. The first part is quite messy, because of the need to deal with determinants. This kind of unpleasant algebra is simplified considerably by using a second-quantised formalism, which is why that was invented.

Evaluate the energy in the form

$$< H >_\Psi = \frac{< \Psi|H|\Psi >}{< \Psi|\Psi >}$$

with the determinantal wavefunction of Eq. (2.93) using an orthonormal set of orbitals ψ_i.
Answer:

$$< H >_\Psi = \sum_i < i|(T + U_{ion}|i > + \frac{1}{2}\sum_{ij}\left[< ij|\frac{e^2}{r_{ij}}|ij > - < ij|\frac{e^2}{r_{ij}}|ji > \delta_{\sigma_i\sigma_j}\right] \ . \tag{2.159}$$

Show that by minimising with respect to the ψ_i^* one obtains the Hartree-Fock equations

$$\begin{aligned}&\left[-\frac{\hbar^2}{2m}\nabla^2 + U_{ion}(\mathbf{r}) + U_{coul}(\mathbf{r})\right]\psi_i(\mathbf{r})\\ - &\sum_j \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')\psi_j(\mathbf{r})\delta_{\sigma_i\sigma_j} = \epsilon_i\psi_i(\mathbf{r})\end{aligned} \tag{2.160}$$

and that the total energy can be written

$$< H >_\Psi = \sum_i \epsilon_i - \frac{1}{2}\sum_{ij}\left[< ij|\frac{e^2}{r_{ij}}|ij> - < ij|\frac{e^2}{r_{ij}}|ji> \delta_{\sigma_i\sigma_j}\right] . \quad (2.161)$$

Qu.2.11 Band structure in the Hartree-Fock approximation

Using Eq. (2.102), calculate the density of states near the Fermi energy to leading order in $(E - E_F)/E_F$. If this result were physically correct what would be the temperature-dependence of the electronic specific heat at low temperature?

The logarithmic singularity you found is not found in better theories than Hartree-Fock. It arises mathematically from the long-range Coulomb interaction (and in particular the k^{-2} divergence) which vanishes when we allow for processes to screen the interaction and make it short-range.

Qu.2.12 Ferromagnetism in the HF approximation

Previously, we considered the unpolarised spin state, which is a paramagnet. Now consider a fully spin polarised state at the same density: the Hartree-Fock Slater determinant corresponds to singly occupying each state in the Fermi sphere. In analogy to Eq. (2.104), compute the total energy of the spin polarised state, and show that this is lower in energy than the unpolarised state if $r_s > 5.45$ in the Hartree-Fock approximation.

The physics here is correct, but the number is very wrong, and the paramagnetic state is believed to be stable up to $r_s \approx 75$. The ferromagnet has larger kinetic energy, because at a fixed density you have always a larger k_F (by a factor of $2^{1/3}$) if only one spin subband is filled. The exchange energy is only between parallel spins and is negative, so prefers spin-aligned states, and the interaction terms become proportionately more important at low densities (large r_s)

Qu.2.13 Thomas-Fermi screening

Check the formulae in Eq. (2.144) and Eq. (2.145). Suppose that the potential $v_{ext} = Q/r$, show that the induced charge density is then of the form

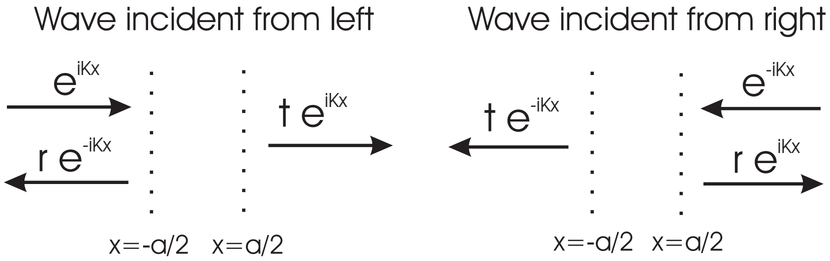
$$\delta n(r) \propto \frac{e^{-r/\xi}}{r}$$

and identify the screening length ξ.

Qu.2.14 * Generalised one-dimensional band theory

Many of the general features of electron bands in a periodic solid can be appreciated without recourse to a detailed model of the lattice potential. In this problem the scattering from the lattice potential is treated in a general way by describing it in terms of wave relfection *and transmission. The algebra in the first two parts is messy - but the answer given in section (b) can be used to answer section (c).*

Consider a 1D solid, lattice constant a, made of "building blocks" ($-a/2 < x < a/2$) that scatter plane waves with a reflection coefficient r and transmission coefficient t ($|r|^2 + |t|^2 = 1$) as shown.



The energy of the plane wave is written as $\epsilon = \hbar^2 K^2/2m$. In the solid, the building blocks are stacked together indefinitely in the x-direction.

(a) Write the solution to the Schrödinger equation in the solid $\psi(x)$, as a linear combination of $\psi_r(x)$ and $\psi_l(x)$ and use Bloch's theorem to relate the wavefunction at each side of the building block (the same theorem applies to the gradient ψ').

$$\psi(x+a) = e^{ika}\psi(x) \quad ; \psi'(x+a) = e^{ika}\psi'(x)$$

Hence, show

$$\cos(ka) = \frac{t^2 - r^2}{2t} e^{iKa} + \frac{1}{2t} e^{-iKa}$$

.

(b) If the transmission coefficient is $t = |t|e^{i\delta}$, it can be shown that $r = \pm i|r|e^{i\delta}$ (it is relatively easy to demonstrate this graphically for the special case when the scattering is localised at $x = 0$ and one can match ψ at the origin). Use this result to eliminate r and show

$$\frac{\cos(ka + \delta)}{|t|} = \cos ka \quad .$$

(c) Since $|t| < 1$, this result shows there are values of K (and hence ϵ) for which no Bloch states exist. Demonstrate this by sketching the left-hand side as a function of K (or preferably ϵ) Use your sketch to illustrate the behaviour for: (i) strong scattering; (ii) weak scattering.

Explain why, in general, electron bands tend to get wider and their gaps narrower as the electron energy increases.

Chapter 3

Electronic structure of selected materials

3.1 Metals

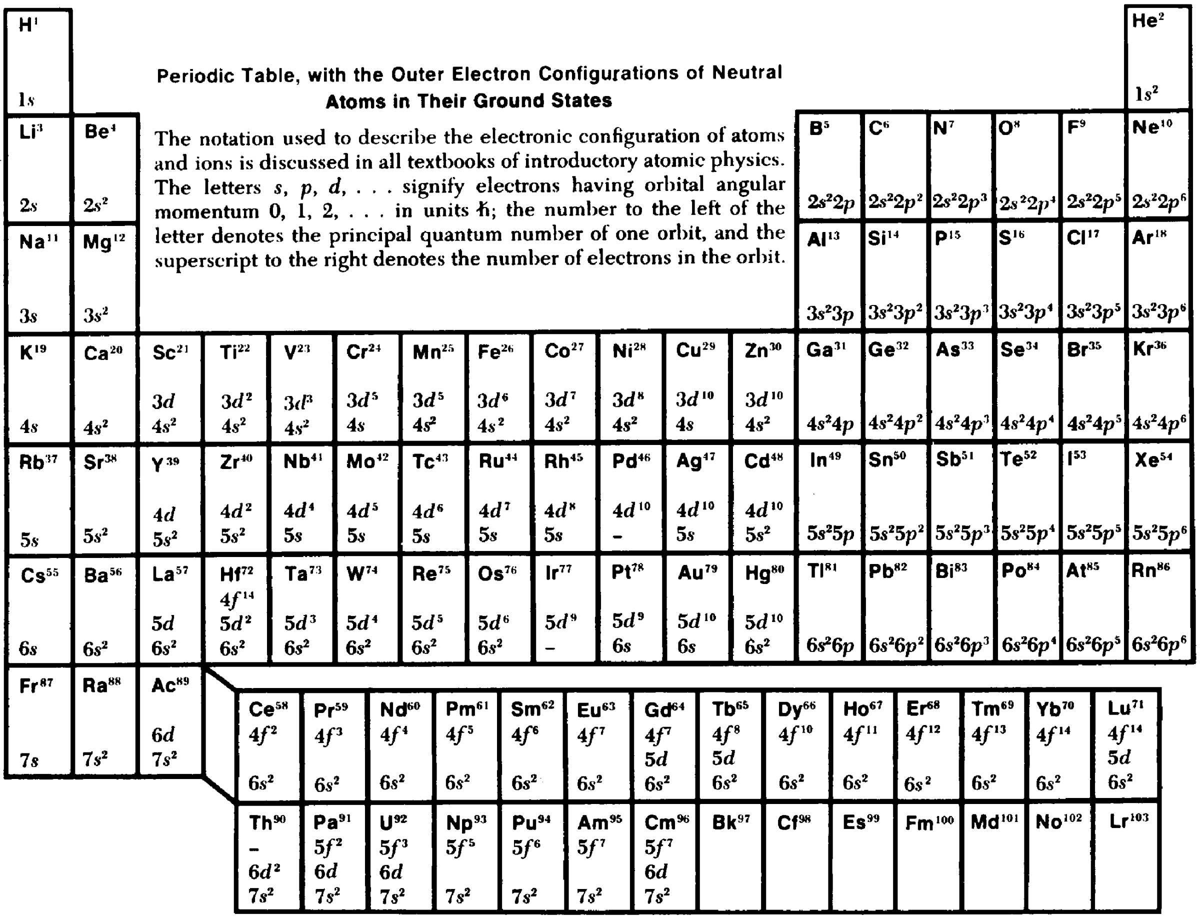
A limited number of elements are well described as nearly free electron compounds. We should expect nearly free electron theory to apply when the electronic density is high, for two major reasons.

The first is that as we saw in the last section, screening of an "external" charge occurs over a range q_{TF}^{-1}, which is short when the density is high. So the nuclear charges are strongly screened, and therefore the ions' scattering potential is weak when they are embedded in a high density electron gas. Secondly, we saw in the homogeneous electron gas that *correlation* effects, which are defined as the effects of electron-electron interactions that cannot be incorporated into interactions only with the *average* effects of all the electrons are weaker at high density. So we should expect the independent electron approximation to work quite well.

The ionic potential itself will be weaker also toward the left had side of the periodic potential. That is because the core electrons are tightly bound, and the valence electrons see only the combined potential of the nuclear charge plus the core. For Na, Mg and Al, for example the 1s, 2s and 2p electrons are bound into a configuration like that of the inert rare gas Ne, and only the 3s and 3p electrons need to be included.

Thus for the alkali metals (row 1 of the periodic table), the fermi surface is alomst spherical and well inside the first Brillouin zone.

In the divalent alkaline earth metals (row 2 of the periodic table) there are two electrons in the outer shell. Consequently the Fermi surface volume must be equal to the volume of the first Brillouin zone (for a solid with one atom per unit cell). Since there is always a gap on the Brillouin zone boundary, these materials must be either insulators, or semi-metals, but because the potential is weak, the gaps are small, and the latter is the



Periodic Table, with the Outer Electron Configurations of Neutral Atoms in Their Ground States

The notation used to describe the electronic configuration of atoms and ions is discussed in all textbooks of introductory atomic physics. The letters s, p, d, . . . signify electrons having orbital angular momentum 0, 1, 2, . . . in units $\hbar$; the number to the left of the letter denotes the principal quantum number of one orbit, and the superscript to the right denotes the number of electrons in the orbit.

H^{1} $1s$																	**He**^{2} $1s^2$
Li^{3} $2s$	**Be**^{4} $2s^2$											**B**^{5} $2s^22p$	**C**^{6} $2s^22p^2$	**N**^{7} $2s^22p^3$	**O**^{8} $2s^22p^4$	**F**^{9} $2s^22p^5$	**Ne**^{10} $2s^22p^6$
Na^{11} $3s$	**Mg**^{12} $3s^2$											**Al**^{13} $3s^23p$	**Si**^{14} $3s^23p^2$	**P**^{15} $3s^23p^3$	**S**^{16} $3s^23p^4$	**Cl**^{17} $3s^23p^5$	**Ar**^{18} $3s^23p^6$
K^{19} $4s$	**Ca**^{20} $4s^2$	**Sc**^{21} $3d$ $4s^2$	**Ti**^{22} $3d^2$ $4s^2$	**V**^{23} $3d^3$ $4s^2$	**Cr**^{24} $3d^5$ $4s$	**Mn**^{25} $3d^5$ $4s^2$	**Fe**^{26} $3d^6$ $4s^2$	**Co**^{27} $3d^7$ $4s^2$	**Ni**^{28} $3d^8$ $4s^2$	**Cu**^{29} $3d^{10}$ $4s$	**Zn**^{30} $3d^{10}$ $4s^2$	**Ga**^{31} $4s^24p$	**Ge**^{32} $4s^24p^2$	**As**^{33} $4s^24p^3$	**Se**^{34} $4s^24p^4$	**Br**^{35} $4s^24p^5$	**Kr**^{36} $4s^24p^6$
Rb^{37} $5s$	**Sr**^{38} $5s^2$	**Y**^{39} $4d$ $5s^2$	**Zr**^{40} $4d^2$ $5s^2$	**Nb**^{41} $4d^4$ $5s$	**Mo**^{42} $4d^5$ $5s$	**Tc**^{43} $4d^6$ $5s$	**Ru**^{44} $4d^7$ $5s$	**Rh**^{45} $4d^8$ $5s$	**Pd**^{46} $4d^{10}$ –	**Ag**^{47} $4d^{10}$ $5s$	**Cd**^{48} $4d^{10}$ $5s^2$	**In**^{49} $5s^25p$	**Sn**^{50} $5s^25p^2$	**Sb**^{51} $5s^25p^3$	**Te**^{52} $5s^25p^4$	**I**^{53} $5s^25p^5$	**Xe**^{54} $5s^25p^6$
Cs^{55} $6s$	**Ba**^{56} $6s^2$	**La**^{57} $5d$ $6s^2$	**Hf**^{72} $4f^{14}$ $5d^2$ $6s^2$	**Ta**^{73} $5d^3$ $6s^2$	**W**^{74} $5d^4$ $6s^2$	**Re**^{75} $5d^5$ $6s^2$	**Os**^{76} $5d^6$ $6s^2$	**Ir**^{77} $5d^9$ –	**Pt**^{78} $5d^9$ $6s$	**Au**^{79} $5d^{10}$ $6s$	**Hg**^{80} $5d^{10}$ $6s^2$	**Tl**^{81} $6s^26p$	**Pb**^{82} $6s^26p^2$	**Bi**^{83} $6s^26p^3$	**Po**^{84} $6s^26p^4$	**At**^{85} $6s^26p^5$	**Rn**^{86} $6s^26p^6$
Fr^{87} $7s$	**Ra**^{88} $7s^2$	**Ac**^{89} $6d$ $7s^2$															
			Ce^{58} $4f^2$ $6s^2$	**Pr**^{59} $4f^3$ $6s^2$	**Nd**^{60} $4f^4$ $6s^2$	**Pm**^{61} $4f^5$ $6s^2$	**Sm**^{62} $4f^6$ $6s^2$	**Eu**^{63} $4f^7$ $6s^2$	**Gd**^{64} $4f^7$ $5d$ $6s^2$	**Tb**^{65} $4f^8$ $5d$ $6s^2$	**Dy**^{66} $4f^{10}$ $6s^2$	**Ho**^{67} $4f^{11}$ $6s^2$	**Er**^{68} $4f^{12}$ $6s^2$	**Tm**^{69} $4f^{13}$ $6s^2$	**Yb**^{70} $4f^{14}$ $6s^2$	**Lu**^{71} $4f^{14}$ $5d$ $6s^2$
			Th^{90} – $6d^2$ $7s^2$	**Pa**^{91} $5f^2$ $6d$ $7s^2$	**U**^{92} $5f^3$ $6d$ $7s^2$	**Np**^{93} $5f^5$ $7s^2$	**Pu**^{94} $5f^6$ $7s^2$	**Am**^{95} $5f^7$ $7s^2$	**Cm**^{96} $5f^7$ $6d$ $7s^2$	**Bk**^{97}	**Cf**^{98}	**Es**^{99}	**Fm**^{100}	**Md**^{101}	**No**^{102}	**Lr**^{103}

Figure 3.1: Periodic table of the elements. From Kittel

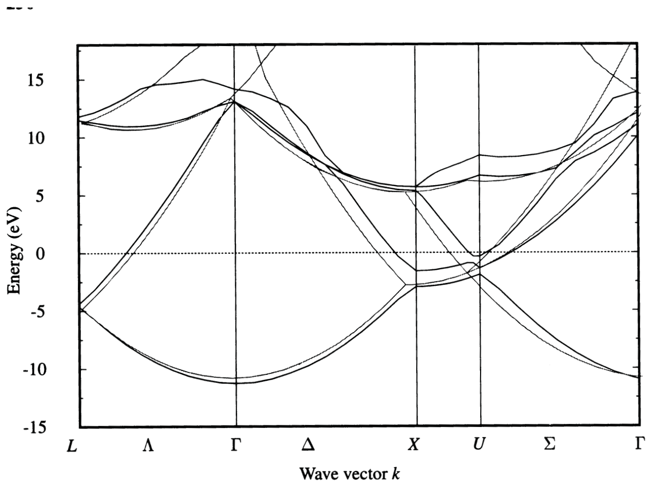


Figure 3.2: Band structure of Al (solid line) compared to the free electron parabolas (dotted line). Calculations from Stumpf and Scheffler, cited by Marder.

case – there is a large electron pocket in the second Brillouin zone, and a counterbalancing hole pocket in the first Brillouin zone. The Fermi surface is now complicated.

Moving to the trivalent metal Al, Fig. 3.2 shows the calculated band structure, compared to the plane wave parabolas, which are remarkably similar.

The tetravalent elements are semiconductors (C,Si,Ge,Sn) or metals toward the bottom of the periodic table (Sn, Pb), with tin existing in two forms. We will discuss these in the next section.

We next turn to metals where the d-shell in the atom plays an important role: the noble metals (Cu,Ag, Au), and the transition metals themselves. The electronic configuration of Cu is $[Ar]3d^{10}4s^1$; for Ag we have $[Kr]4d^{10}5s^1$ and Au $[Xe]5d^{10}6s^1$. It turns out that the d-shell electrons are not so tightly bound to the atom that they can be considered as part of the core: however the $3d$ wavefunctions are more tightly bound than the $4s$, which means that in a tight-binding picture, the overlap between the $3d$-states on neighbouring atoms will be smaller than the overlap between neighbouring $4s$-states. We would expect to see a broad s-band(holding two electron states per atom), and a narrow d-band (holding 10 electron states per atom), and this is in fact the case (see Fig. 3.3). However, these sets of band overlap, and in the crystal, the atomic d- and s-states are not eigenstates, of course. So where the bands cross, they mix, and this *hybridization* is very important for the associated magnetic phenomena. Another very important consequence of the hybridization in Cu is that the metal has an open Fermi surface (Fig. 3.4): in the $< 100 >$ direction, the rapidly dispersing

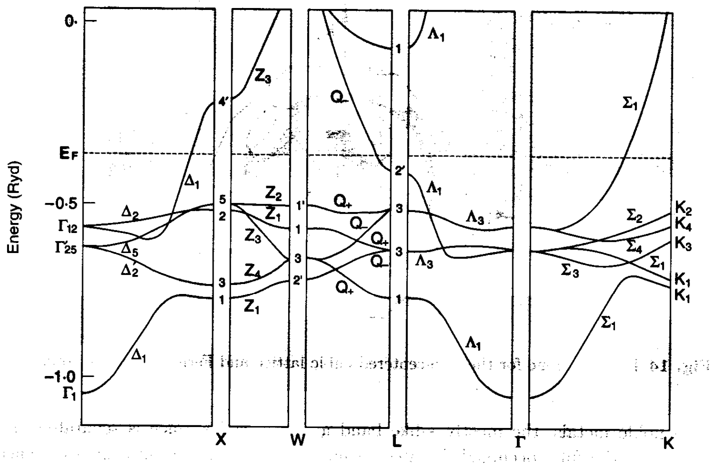


Figure 3.3: Band structure of Cu metal [from G.A.Burdick, *Phys. Rev.* **129**,138 (1963)], cited by Grosso and Parravicini

s-like band crosses the Fermi surface, but there is no crossing in the 111 direction. This topology has important consequences for transport properties like magnetoresistance.

In transition metals (partially filled 3d, 4d and 5d shells) or rare earth metals (partially filled 4f or 5f shells), the electronic states are determined by a mixing of the quasi-localised d- or f- states with the itinerant s- or p- states. However, in many of these metals (and especially in their compounds), the bands are narrow enough that the independent electron approximation is no longer even approximately valid. This is the realm of *strongly-correlated* electronic systems, and we shall need different tools to understand them.

3.2 Semiconductors

The group IV elements, diamond, silicon, germanium, and one form of tin (gray-Sn) crystallise in the diamond structure, which is f.c.c. with two atoms per unit cell.

The electronic configurations are just $[\ ]ns^2np^2$ with n the principal quantum number. The band structures of Si, Ge, and GaAs are shown in Fig. 3.6 and Fig. 3.7. The maximum of the valence bands of all the materials is at Γ. Si and Ge are both *indirect* gap materials, because the conduction bands have minima either in the (100) direction (Si) or the (111) direction (Ge).

In contrast to the metals, these *covalently*-bonded solids are much less dense - with only four nearest neighbours at the corners of a tetrahedron - rather than the twelve of a close-packed metal. These materials have forgone the strong cohesion that can be gained in a high-density metallic state for

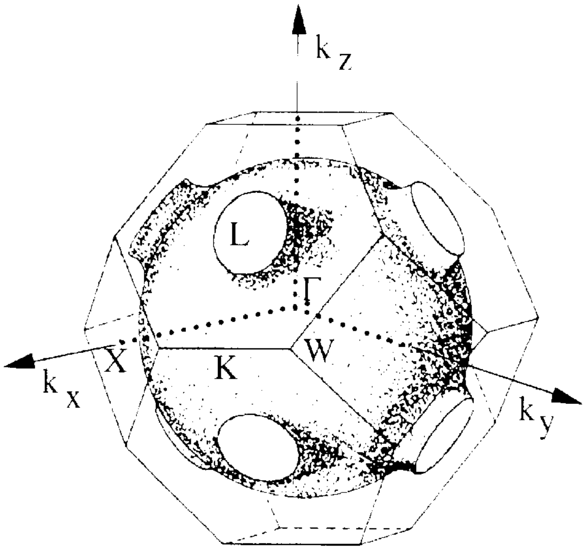


Figure 3.4: Fermi surface of Cu

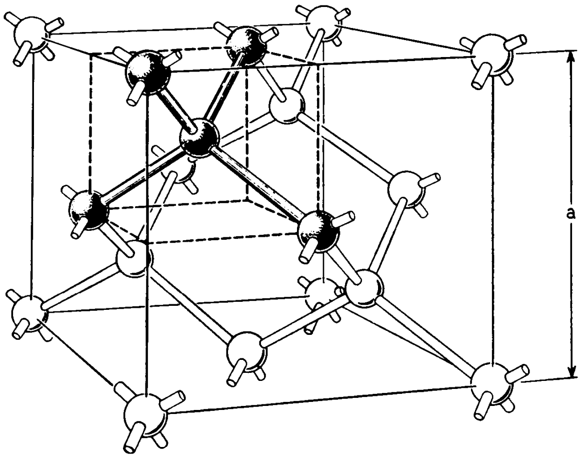


Figure 3.5: Tetrahedral bonding in the diamond structure. The zincblende structure is the same but with two different atoms per unit cell

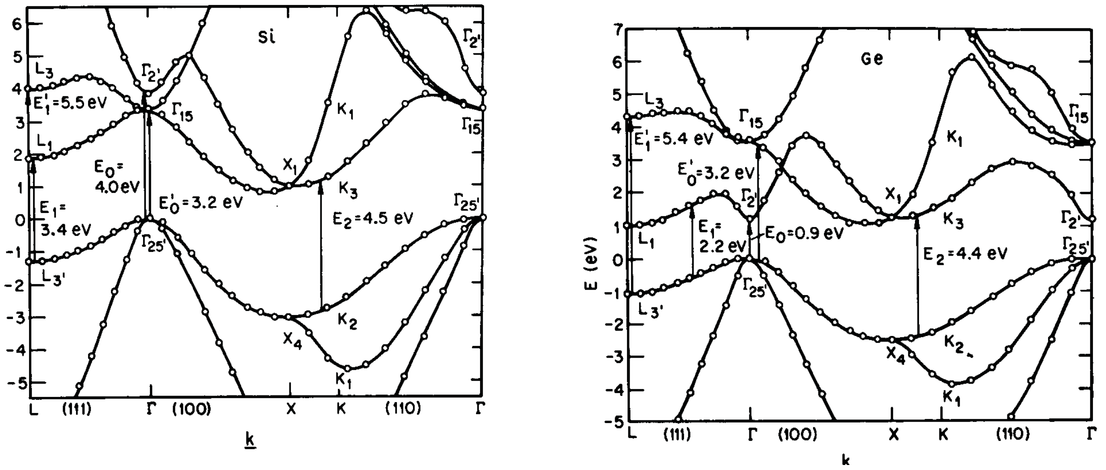


Figure 3.6: Pseudopotential band structure of Si and Ge [M.L.Cohen and T.K.Bergstresser *Phys.Rev***141**, 789 (1966)]. The energies of the optical transitions are taken from experiment.

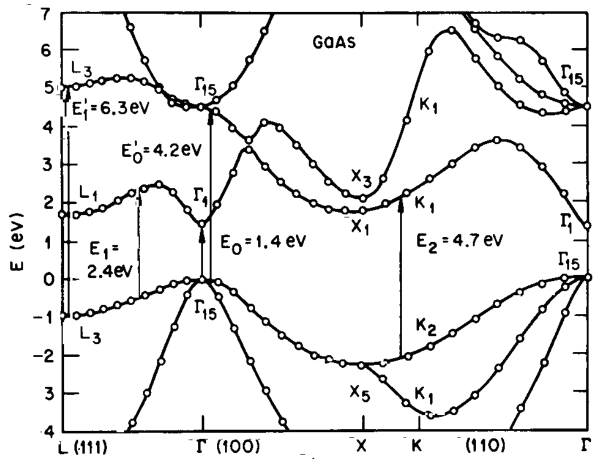


Figure 3.7: Band structure of GaAs [M.L.Cohen and T.K.Bergstresser *Phys.Rev***141**, 789 (1966)]

the directed bonding in a covalent solid. With the electronic configurations above we have 4 electrons per atom (8 per unit cell) to accommodate into eight bands formed from the s-p hybridised orbitals (one s and three p per atom). We need to fill half the available bands. When this is done in Si, we see from the figure that we fill the states up to a substantial gap on the surface of the Brillouin zone. The valence bands are separated from the conduction bands by a large amount, and we need to understand where this overall energy gap comes from.

The tetrahedral directions are well known in the theory of chemical bonds. We know that we can combine the 1s and 2p orbitals to make a new basis set of sp^3 hybrid orbitals, which point in the $< 111 >$ directions. These orbitals have the nice feature that they are directional: the lobes point mostly in the direction of the neighbouring atom.

Within the tight-binding description, we would then consider principally terms in the Hamiltonian that make overlaps to the neighbouring atom with the orbital pointing in the same direction; if one neglects all other overlap terms, then clearly the Hamiltonian factors into a product of *non-interacting* molecules. The band-structure will then reflect these (2x2) subunits: the mixing of a pair of degenerate orbitals will yield a split bonding and antibonding combination of orbitals. There are then four occupied bonding orbitals, and four unoccupied antibonding orbitals, with no dispersion.[1] We have found no dispersion so far because we neglected two things: one is the energy cost of (in chemical language) “promoting” an s-electron into a p-state; the other is the further hybridisation or “hopping” terms that we neglected. All this leads to broadening of the bonding and antibonding levels into bands; but the fundamental origin of the gap can be seen by chemical means.

It is also instructive to look at this problem from the nearly-free-electron point of view. Except close to the chemical potential, most of the band dispersion is captured by the free electron model. Near the chemical potential, there is a substantial gap. Because the gap occurs after four bands are filled, we need to look at the band gaps on the fourth Brillouin zone boundary. This is drawn in Fig. 3.8 for the diamond lattice in the extended zone scheme.

Bands near the zone centre

It is of interest to look at the bands near the Γ-point, because then some symmetry principles help the analysis. Since the zincblende structure has inversion symmetry about each atom, and since at $k = 0$ the wavefunction is the same in each unit cell, the states can be separated into those that are even or odd about any lattice site (for our purposes s and p). Additionally, if we have the diamond lattice there is a further symmetry point midway between the two atoms in the unit cell; the states can also be classified as odd/even about

---

[1] Note that we would have derived the same electronic structure even if the solid was amorphous – provided solely that each atom has four neighbours in approximately the correct orientation

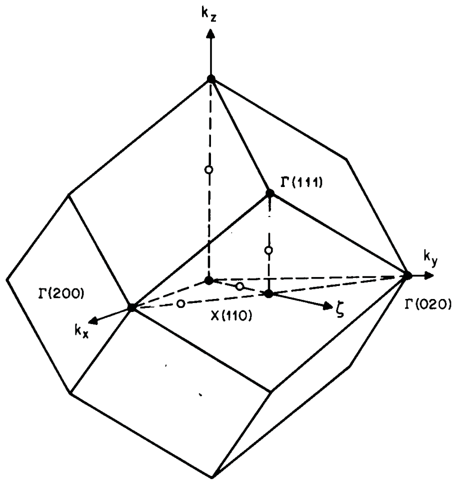


Figure 3.8: The fourth Brillouin zone (or Jones zone) for a face-centered cubic crystal. The surface consists of $\{110\}$ planes. Points on the surface of the Brillouin zone are labelled with both their conventional notation and coordinates in units of $2\pi/a$, where a is the cubic lattice constant

this point. So we must have the following sets of states:

bonding (antibonding) $s_1 \pm s_2$
bonding (antibonding) $p_1 \pm p_2$

(the subscript refers to the two atoms in the cell) The $p-$ states are triply degenerate (in the absence of spin-orbit coupling which we will come back to in a moment) because the crystal is cubic – this degeneracy is analogous to the orbital degeneracy of p_x,p_y and p_z states in an atom. We expect these states to be ordered in the way shown in Fig. 3.9.

In zincblende materials, the admixture of states will reflect the partial ionicity, as noted in the figure: if atom 2 is the more electronegative atom, $\gamma^{(\prime)} > 1$, but the bonding/antibonding character is essentially preserved.

Now let us include spin-orbit interactions, which will come with a term $\lambda \mathbf{L} \cdot \mathbf{S}$ in the Hamiltonian. This will split the p-states in Fig. 3.9, and we can easily see how. Without spin orbit, we had a sixfold-degenerate level altogether – 3 (orbital) x 2 (spin). After coupling, the states must be eigenfunctions of the total angular momentum J, which can take on the values $J = 3/2$ (fourfold degenerate), and $J = 1/2$ (twofold degenerate). We end up with the bands show in Fig. 3.10 and in particular with the characteristic "light" and "heavy" mass bands for holes.

Semiconducting band gap in the nearly-free-electron picture

In the nearly free electron picture, we should then expect a gap equal to $2U(220)$, because we can scatter from k to the degenerate $k-(220)$. Of course, there are other ways to scatter by the same reciprocal lattice vector, and in particular we note that $(220) = (111) + (11\bar{1})$, so we can get the same effect by scattering twice from the lattice potential $U(111)$. Since $|(111)| < |(220)|$, $U(111)$ is the larger, and its effect in second order can overcome the

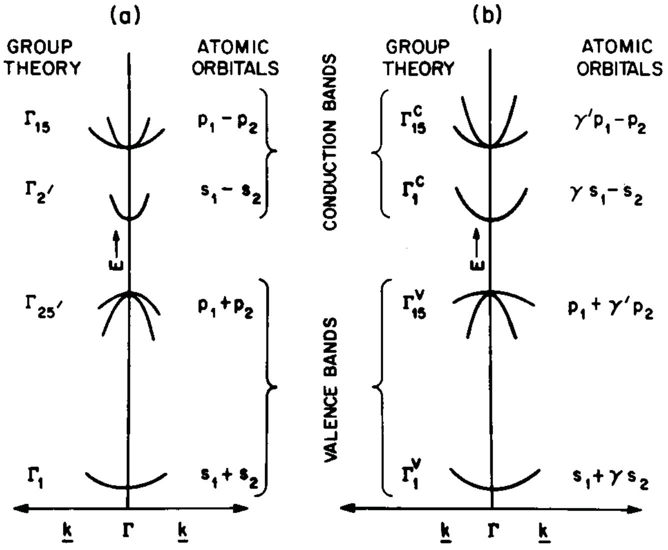


Figure 3.9: Valence and conduction bands near k=0 according to a tight-binding prescription. (a) is for diamond crystals, and (b) for zincblende; spin-orbit splitting is neglected. (From Phillips.)

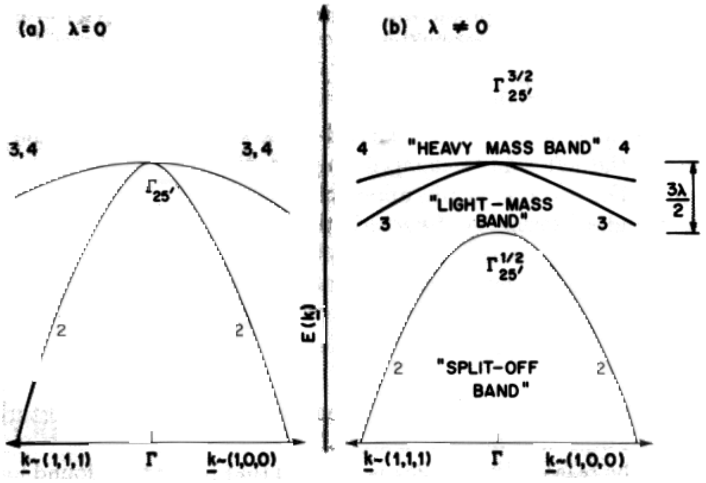


Figure 3.10: Sketch of the valence bands of diamond or zincblende structure semiconductors showing the effect of spin orbit interaction on the topmost valence bands

effect of $U(220)$ in first order [2]

We can then guess that the effective scattering potential might be something like

$$U_{eff}(220) = U(220) + \frac{U(111) \times U(11\bar{1})}{\Delta E} \ , \tag{3.1}$$

where we can staightforwardly see how this formula comes from second-order perturbation theory, with

$$\Delta E = \frac{\hbar^2}{2m}\left[|(110)|^2 - |(110) - (111)|^2\right] = \frac{\hbar^2}{2m}\left(\frac{2\pi}{a}\right)^2 \ . \tag{3.2}$$

In the zincblende structure, with two different atoms A and B, one has from Eq. (2.31)

$$U(111) = -U(11\bar{1})^* = [U^A(111) + U^B(111)] + i[U^A(111) - U^B(111)] \ , \tag{3.3}$$

and so we have a gap which adds the symmetric ("covalent") and antisymmetric ("ionic") components in quadrature just like the one dimensional example we studied in Problem 2.6.[3]

Table 3.11 shows the average band gap for a variety of tetrahedrally coordinated crystals, separated into the covalent and ionic parts.[4] There are two very compelling trends that are quite visible in the data on the average band gap. One is that the band gap decreases as one moves down the periodic table. This we understand most easily by reference to Fig. 3.9, which shows that at least at $k = 0$ the hybrid sp^3 states are not the right description, and of course in a real atom, the energy of the s-state is lower than the p : $E_{ns} < E_{np}$. This comes about because the s-electrons penetrate closer to the atomic core than the corresponding p-state, so they see a less well-screened potential and lie deeper. This separation accentuates with increasing row number in the periodic table, and destabilises the formation of sp^3 hybrids. So elemental Sn is found in two forms, the diamond structure gray tin, and the metallic white tin. The next element down in the groupIV is lead, which is always found as a close-packed metal.

Gray-Sn is an interesting case, because although the average band gap is still positive, just at the gamma point the antibonding s-states fall below the bonding p-levels (a trend that one can already see from looking at the band structures of Si and Ge above), so it is then a zero-gap semiconductor, because the lowest conduction band and the highest valence band are now degenerate by symmetry.

The second trend is that the average band gap increases with ionicity. In each of the sequences Si:AlP:MgS and Ge:GaAs:ZnSe the band gap increases as the ionicity grows; but since each sequence is taken from the same row of the periodic table, the average symmetric potential is presumably not strongly changed. Eventually, of course, if the ionicity is large enough, one expects (and finds) that the fourfold-coordinated zincblende or wurtzite structures are unstable to crystal structures such as rocksalt or caesium chloride, where the Coulomb (Madelung energy) is greatest.

3.3 Narrow band and strongly correlated systems

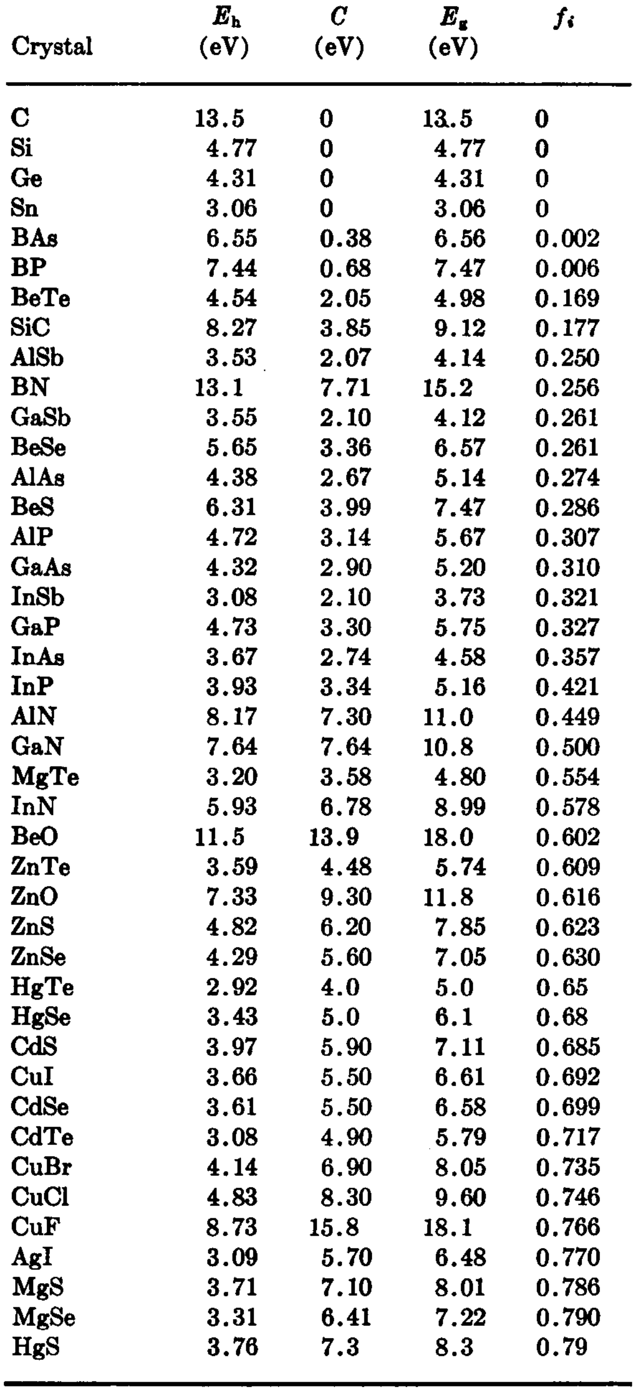
Strongly correlated systems we define loosely as those which have an interaction energy between the electrons that is comparable or larger to their

---

[2] It turns out that often the (220) vector comes very near a node in the pseudopotential, so $U(220)$ is anomalously small

[3] This approach has been carefully worked out by J.C.Phillips, *Bonds and Bands in Semiconductors*, AP, New York 1972

[4] This comes from an empirical analysis, rather than from a direct pseudopotential calculation



Crystal	E_h (eV)	C (eV)	E_g (eV)	f_i
C	13.5	0	13.5	0
Si	4.77	0	4.77	0
Ge	4.31	0	4.31	0
Sn	3.06	0	3.06	0
BAs	6.55	0.38	6.56	0.002
BP	7.44	0.68	7.47	0.006
BeTe	4.54	2.05	4.98	0.169
SiC	8.27	3.85	9.12	0.177
AlSb	3.53	2.07	4.14	0.250
BN	13.1	7.71	15.2	0.256
GaSb	3.55	2.10	4.12	0.261
BeSe	5.65	3.36	6.57	0.261
AlAs	4.38	2.67	5.14	0.274
BeS	6.31	3.99	7.47	0.286
AlP	4.72	3.14	5.67	0.307
GaAs	4.32	2.90	5.20	0.310
InSb	3.08	2.10	3.73	0.321
GaP	4.73	3.30	5.75	0.327
InAs	3.67	2.74	4.58	0.357
InP	3.93	3.34	5.16	0.421
AlN	8.17	7.30	11.0	0.449
GaN	7.64	7.64	10.8	0.500
MgTe	3.20	3.58	4.80	0.554
InN	5.93	6.78	8.99	0.578
BeO	11.5	13.9	18.0	0.602
ZnTe	3.59	4.48	5.74	0.609
ZnO	7.33	9.30	11.8	0.616
ZnS	4.82	6.20	7.85	0.623
ZnSe	4.29	5.60	7.05	0.630
HgTe	2.92	4.0	5.0	0.65
HgSe	3.43	5.0	6.1	0.68
CdS	3.97	5.90	7.11	0.685
CuI	3.66	5.50	6.61	0.692
CdSe	3.61	5.50	6.58	0.699
CdTe	3.08	4.90	5.79	0.717
CuBr	4.14	6.90	8.05	0.735
CuCl	4.83	8.30	9.60	0.746
CuF	8.73	15.8	18.1	0.766
AgI	3.09	5.70	6.48	0.770
MgS	3.71	7.10	8.01	0.786
MgSe	3.31	6.41	7.22	0.790
HgS	3.76	7.3	8.3	0.79

Figure 3.11: Average energy gaps in binary tetrahedrally coordinated crystals, separated into a covalent part E_h, an ionic component C, and a total gap $E_g = \sqrt{E_h^2 + C^2}$. The "ionicity" is defined as $f_i = C^2/E_g^2$. From Phillips [1973]

kinetic energy of motion through the solid. Notice that this definition of kinetic energy corresponds to the bandwidth of the relevant electronic states, not the total kinetic energy, which includes terms from localisation of electronic states in the nuclear potential. In practice, we shall use it to mean systems where the approximate mean-field treatment of the Coulomb interaction between electrons is unreliable, or fails entirely.

Correlations are responsible for many physical phenomena in solids, including magnetism, localisation and metal-insulator transitions, charge- and spin-density waves, high-temperature superconductivity and the fractional quantised Hall effect.

3.3.1 Interactions and the "Hubbard U"

Presupposing the answer, we should really be discussing the electronic structure in terms of a localised tight-binding description, and let us for a moment consider a model system that has one orbital per site, and one electron per site. Band theory will tell us that we have a half filled band, and therefore a metal. The occupied states in band-theory are indexed by momentum, of the form

$$\psi_{k\sigma} = N^{-\frac{1}{2}} \sum_{i=1}^{N} e^{i\mathbf{k}\cdot\mathbf{R}_i} \phi_\sigma(\mathbf{r} - \mathbf{R}_i) \tag{3.4}$$

Note that each electron state is spread out throughout the whole crystal: we get an average occupancy of 1 per site by adding together N terms (one from each occupied k-state), each of which has amplitude $1/N$. The dominant repulsive term in the Coulomb energy comes from when the electrons lie on the same site, and usually this goes by the name of the Hubbard "U". Formally it is

$$U = \int\int d\mathbf{r}\, d\mathbf{r}'\, |\phi(\mathbf{r})|^2 |\phi(\mathbf{r}')|^2 \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \ . \tag{3.5}$$

The contribution of the interaction term to the energy just depends on the probability of *double* occupancy; in the extended state a little thought will convince you that the probability of double occupancy is $\frac{1}{2}$.

The kinetic terms and the interaction terms work against each other: in a Hartree-Fock state of extended orbitals, the kinetic energy goes down because of overlap (by of order t, the hopping matrix element), but the Coulomb energy goes up by $O(U)$. If $U \gg t$, one might guess that this is not a good starting point – in this limit it would seem to be better to localise exactly one electron on each site and forbid hopping between sites – a state which has total energy 0, because neither the kinetic nor the interaction terms contribute to the energy. We will come back to this picture of a Mott metal-insulator transition in a subsequent section; for the purposes of our survey of electronic structure, it simply causes us to consider the values and trends in interaction strengths relative to the electronic bandwidths.

In our simplified atom, we can define the energies of singly and doubly occupied states as follows

$$H_{atom} = \epsilon_o \sum_\sigma n_\sigma + U_{atom} n_\sigma n_{-\sigma} \tag{3.6}$$

where σ is the spin, and n the number (operator). So the energy for zero, single and double occupancy is

$$\begin{aligned} E(0) &= 0 \\ E(1) &= \epsilon_o \qquad (3.7) \\ E(2) &= 2\epsilon_o + U \end{aligned}$$

$$\tag{3.8}$$

so that if we know the energies of the atomic levels, we will define

$$U_{atom} = E(2) + E(0) - 2E(1)\,. \tag{3.9}$$

Suppose, as is often the case, that double occupancy corresponds to the neutral atom; then $E(2) - E(1) = I_1$, the first ionisation potential, and $E(1) - E(0) = I_2$, the second ionisation potential. Then in terms of measurable atomic quantities,

$$U_{atom} = I_1 - I_2\ . \tag{3.10}$$

The second ionisation potential is generally larger than the first, because the ion is increasingly charged; so we get a U_{atom} (defined this way) which is *negative*, based on the "0" state as empty of electrons. But of course we have really defined a *positive* U for *holes*, which are the elementary excitations of removing charge from the neutral atom. Notice that the definition of U, even in the atom, will depend on which configuration we choose as the “ground” state.

3.3.2 3d Transition Metals

In the 3d transition metals (from Sc to Zn) the physical properties are principally determined by the filling of the d-shell, which has a formal configuration ranging from $3d^1 4s^2$ (Sc) to $3d^{10} 4s^2$ (Zn). The inner shells are quite localised and play little role in the low energy or cohesive properties of the solid. Note that the $4s$ orbitals are filled before the d-shell, in the atom. The $4s$ wavefunction has a large amplitude at the core, and sees a higher effective charge on the core because it penetrates inside the inner shells. However, the $4s$ also extend out further from the nucleus than the $3d$ states. The $4s$ orbital is pushed out to help maintain orthogonality with the s-states in the core, and has 3 radial nodes; since are no d states in the core this allows the d-orbitals to be more tightly bound. Because the d-states are more localised

	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu
E_1	6.5	6.8	6.7	6.8	7.4	7.9	7.9	7.6	7.7
E_2	12.8	13.6	14.7	16.5	15.6	16.2	17.1	18.2	20.3
$-E_c$	1.2	2.8	3.1	4.0	4.6	5.7	6.4	6.7	
$-E_F$	3.5	4.3	4.3	4.5	4.1	4.5	5.0	5.1	

Table 3.1: First (E_1) and second (E_2) ionisation potentials, position of the center of the d-band (E_c) and Fermi energy (E_F) for the 3d transition metals (in eV).

the effects of Coulomb repulsion are more important. This can already be seen in the atomic configurations in that Cr $3d^5 4s^1$ and Cu $3d^{10} 4s^1$ prefer to depopulate an s-state in order to gain the stabilization of a half-filled or filled shell. Another trend is in the ionization potentials: both the first and second ionization potentials increase as the d-shell is filled. This arises because the d-electrons screen imperfectly the (increasing) nuclear charge for an added d-electron in their midst.

In the solid, the s-orbitals will hybridise most strongly with their neighbors and produce a wide band. The d-orbitals overlap much less strongly, but this narrow band of d-states lies in the midst of the broad s-band and hybridises with it. Some of the atomic trends persist in that the increasing ionization potential of the ions is mirrored in a lowering of the d-bands away from the chemical potential.(See Table 3.1) The net result is that the center of the d-band in the metal moves from about $-1eV$ in Sc to $-7eV$ in Ni, while the Fermi energy falls by less – $-3.5eV$ (Sc) to $-5.5eV$ (Ni) as the d-band is filled. Overall, the d-band width is in the range of $5 - 8eV$.

In oxides, the effect of interactions will be stronger because the ionicity will mean that the transition metal ion is in a M^{2+} or M^{+} state, and therefore the relevant atomic excitation which fixes the band center is the second ionization potential. As can be seen in Table 3.1, this increases more rapidly than the first ionization potential. Notice that the atomic U_{atom} is smallest on the left hand side of the d-series, and increases to the right.

Because of the narrow band-width, the Coulomb repulsion effects cannot be ignored. Indeed, it was pointed out long ago that the insulating character of some transition metal compounds cannot be understood from conventional band theory. This problem was addressed by Mott , Anderson, and Hubbard [5], who showed that if the d-d Coulomb interaction is larger than the band-width a correlation gap will form with the Fermi level in the gap.

For real d-levels, instead of just the simple single orbital case described above, we need to consider more interaction terms. The correlations between electrons on the same ion are dominated by the effects of the exchange

---

[5] See N.Mott, *Metal Insulator Transitions*

interactions

$$J^{ij}_{mm'} = \langle \psi_{im}(r_1)\psi_{jm'}(r_2)|\frac{1}{r_{12}}|\psi_{jm'}(r_1)\psi_{im}(r_2)\rangle \tag{3.11}$$

where ψ is a localized d orbital around an atom, i, j denotes the sites and m, m' are the l_z quantum numbers. The atomic wave functions are extremely localized and decay exponentially fast away from the atom, so the interatomic terms are much smaller than the *intra*-atomic terms ($i = j$). If the transfer integral decays as e^{-qr}, then the interatomic exchange will decay as e^{-2qr}, and can generally be neglected.

Estimated magnitudes of the interactions in Eq. (3.11) are $U = J^{ii}_{mm} \approx 20eV$, and $J = J^{ii}_{m\neq m'} \approx 1eV$. The large disparity in magnitude between U and J encourages ferromagnetic alignment of the spins on the same atom so that the Pauli principle avoids the repulsion U. This is Hund's rule, and explains why d-metals typically have large magnetic moments.

One can define an average exchange energy $\bar{U}$ per pair of electrons. This gives the energy gained from switching from antiparallel to parallel spins, on the assumption that different angular momentum states are occupied equally. This gives an estimate for the interaction in the Hartree-Fock scheme as $\bar{U}_{HF} \approx (U + 4J)/5 \approx 5\ eV$. Although we have used the same notation as for U_{atom} (and we will use this U in the same way as in Eq. (3.6)), it is defined here in a different way: this is for transitions on a *neutral* atom. We consider double occupancy in a single orbital, but counterbalanced by zero occupancy in an orthogonal d-orbital.

3.3.3 Interatomic screening

The *effective interaction* $\bar{U}$ we defined above provides an example of intra-atomic screening: charge fluctuations between orbitals in a single atom, where the atom as a whole is kept neutral. Such an interaction in the metal would split the d-band into separate (Hubbard) bands, because the bandwidth $\approx 5\ eV$ is of the same order. However, in many situations the effective U is much smaller, because of screening effects by electrons in other bands, especially the $4s$.

To see how this works in practice, let's look carefully at a particular case relevant for the strongly correlated d-systems with nearly closed d-shells: Cu^{2+} or Ni^{+}, which are formally close to d^9s^0. We can estimate an "atomic" U for the Cu^{2+} or Ni^{+} configurations from the energy of the transition

$$2(3d^9) \rightarrow 3d^{10} + 3d^8. \tag{3.12}$$

which is also the difference between the first and second "ionization potentials" of the $3d^{10}4s^0$ configuration (see Table 3.2). Defined this way, if no s-states are occupied, the energy cost is about 13.5 eV. Most of this is the isotropic Coulomb cost of a doubly-charged (versus two singly-charged) ions. In a solid, this is an overestimate of the repulsion, because if the s-electrons are allowed to relax to compensate for the charging of the ions, we should consider the transition

$$2(3d^94s) \rightarrow 3d^84s^2 + 3d^{10} \tag{3.13}$$

which costs only 1.8 eV for ground states of the neutral atom (see Table 3.2).

d^8s^2	d^9s	$d^{10}s^0$	d^9s^0	d^8s^0
	Ni		Ni^+	Ni^{++}
0	0.025	1.83	7.63	25.78

Table 3.2: Term values of the Ni atom in various stages of ionization

In transition metals, this screening is quite efficient, and the effective values of the interaction energy are of order 2-3 eV, less than the band width, so that the transition metals are indeed metallic. The Coulomb exchange interactions are of course crucial in understanding the magnetic properties, but they do not destroy the one-electron picture of the bandstructure.

An additional effect of the solid environment is the crystal-field splitting of the d-levels, which breaks the orbital degeneracy. In the metals, the crystal field splitting is generally smaller than the overall d-band width and the exchange energy J, and is not important. However, if the crystal-field splitting is large (as in the oxides, for example) the filling of the orbitals will proceed independently of the spin - and generally the configurations will have low spin.

3.3.4 Transition metal oxides and the Mott transition

In the oxides, providing the O-p states lie well below the chemical potential, there is complete charge transfer of the metal s-electrons onto the oxygen. The screening by the s-electrons disappears, and only the relatively ineffective screening by the intra-atomic d-levels persists. Consequently Ni is a metal (good screening) whereas NiO is an insulator, despite the band theory prediction.

V_2O_3 is the prototype for metal-insulator transitions in the oxides. It marks the boundary in the transition metal series between metals and insulators; because U_{eff} is not too large, a metal to insulator transition is found at moderate temperatures and pressures, where the conductivity changes abruptly by several orders of magnitude. To the right of Vanadium, the insulators dominate, with NiO the classic example. As we discussed above, this crossover is occurring because the d-band is moving lower in energy and intra-atomic screening by the s-band is becoming less effective. Even here, the simple theory is not complete, because the transition in V_2O_3 is first-order, and accompanied by a large volume change. The change in screening at the transition changes the interatomic interactions between ions, and feeds back to a change in the crystalline structure.

Questions on chapter 3

Qu.3.1 Graphite A single sheet of graphite has two carbon atoms in the unit cell at positions $\mathbf{d}_1 = 0$ and $\mathbf{d}_2 = (a/\sqrt{3})(0, 1, 0)$. The translation vectors for the two-dimensional hexagonal lattice are $\mathbf{t}_1 = (a/2)(1, \sqrt{3}, 0)$ and $\mathbf{t}_1 = (a/2)(-1, \sqrt{3}, 0)$.
The electronic configuration of the carbon atom is $1s^2 2s^2 2p^2$, and ignoring the $1s$ core states, we need to make a band structure from the s, p_x, p_y abd p_z orbitals. Because s, p_x and p_y orbitals are even under reflection through the plane, and p_z odd, the two sets do not mix. The first three states hybridise to form $\sigma-$bonds with a large gap between the bonding and anti-bonding orbitals; we consider here the π-bonds from the p_z.

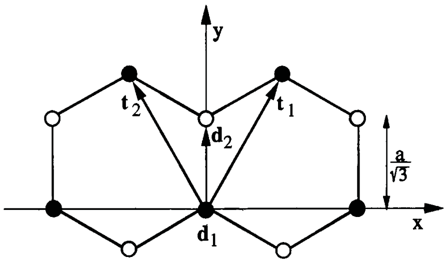


Figure 3.12: Two dimensional structure of graphite

(a) Construct Bloch states that consist of a linear mixture of the two p_z orbitals in the unit cell, and show how this gives rise to the secular equation

$$\begin{vmatrix} E_p - E & V(pp\pi)F(\mathbf{k}) \\ V(pp\pi)F^*(\mathbf{k}) & E_p - E \end{vmatrix} = 0 \ , \tag{3.14}$$

where $v(pp\pi)$ is the two center integral between neighbouring p_z orbitals, and

$$F(\mathbf{k}) = 1 + 2\cos\left(\frac{k_x a}{2}\right)\exp\left(-i\frac{\sqrt{3}k_y a}{2}\right) \ . \tag{3.15}$$

(b) Draw the first Brillouin zone for the 2D hexagonal lattice. Determine a formula for the dispersion curves for the two eigenstates, and plot them in the directions ΓP, and ΓQ. (Here $\Gamma = (000)$, $P = (2\pi/a)(2/3, 0, 0)$, $Q = (2\pi/a)(1/2, 1/2\sqrt{3}, 0)$.

(c) Where will the π-bands lie in energy relative to the sp^2 σ- orbitals? Is graphite a metal or an insulator?

(d) Carbon nanotubes are formed by curling a graphite sheet into a tube, connecting the atoms with periodic boundary conditions. There are many ways to do this, and the different nanotubes can be indexed by the vector $m\mathbf{t}_1 + n\mathbf{t}_2$ that is the new period. Ignoring the effects of curvature of the planes, but including the new periodicity, draw the new superlattice

Brillouin zone boundaries on top of the 2D hexagonal lattice Brillouin zone, in the cases: m=10, n=0; m=10, n=1; m=10, n=2. Discuss the effects of this new periodicity on the electrical resistance of a carbon nanotube.

Qu.3.2

Describe the role of Bloch's theorem in determining the form of electronic wavefunctions in a periodic solid.

A one-dimensional solid lies in the x-direction. Sketch the Bloch wavefunctions at the centre and edge of the first Brillouin zone of a band composed of:
(a) atomic s states, and
(b) atomic p_x states.

Sketch typical energy curves in the reduced zone scheme for:
(a) an alkali metal;
(b) a divalent metal;
(c) a non-magnetic 3d transition metal;
(d) a direct band-gap semiconductor;
(e) an indirect band-gap semiconductor;
(f) a wide-gap insulator.
Comment on the physical properties of the materials that can be related to your sketches.

Qu.3.3 Band structure of d-band metals

In many transition metals a narrow d-band lies within a broad energy band originating from $s-$orbitals. This question discusses the band structure using a simple one-dimensional model contructed from a tight-binding Hamiltonian with one s-orbital $\phi_s(r)$ and one d-orbital $\phi_d(r)$ per atom; the atoms are arranged in a linear chain of lattice constant a.

Write down two Bloch states $\phi_s(k)$ and $\phi_d(k)$ formed from the atomic s- and d- states respectively. The eigenstates must be linear combinations of these. Hence show that the one-particle bandstructure $E(k)$ can be determined from the determinantal equation

$$\begin{vmatrix} E_s - 2t_{ss}\cos(ka) - E(k) & -2t_{sd}\cos(ka) \\ -2t_{sd}\cos(ka) & E_d - 2t_{dd}\cos(ka) - E(k) \end{vmatrix} = 0 \quad .$$

Identify and explain the parameters appearing in the determinantal equation, and discuss the approximations made that lead to this form.

Discuss why you would expect that $t_{ss} > |t_{sd}| > t_{dd}$.

Sketch the dispersion $E(k)$, in the cases (i) $E_d \ll E_s - 2|t_{ss}|$, and (ii) $|E_d - E_s| < -2|t_{ss}|$.

Discuss the relevance of this model to the electronic bandstructure of Cu metal.

Qu.3.4

(a) Discuss the trends in electronic bandstructure and cohesion in the group IV elements.

*(b) Ni, NiS, and NiO all have crystal structures that would lead to an odd number of electrons per unit cell. Explain why Ni and NiS are metals, but NiO is an insulator.

Chapter 4

Collective phenomena

4.1 Response functions and collective modes

Many experimental measurements in solids take the form of measuring the linear response to a probe field. Familiar examples are the electrical conductivity $j = \sigma E$, the dielectric function $D = \epsilon\epsilon_o E$, and the magnetic susceptibility $M = \chi H$. These quantities (σ, ϵ, χ) appear in simple models of electromagnetism just as material constants; in general they are however not just numbers, but functions of space and time.

A generalised response function measures the response of the system at a point r and time t to a field applied at some other point r' at a time t'. They can all be described by the formal relationship

$$u_\alpha(\mathbf{r},t) = \int d\mathbf{r}' \int dt' \chi_{\alpha\beta}(\mathbf{r},t;\mathbf{r}',t') F_\beta(\mathbf{r}',t') \ , \tag{4.1}$$

where $\mathbf{u}$ is the response, $\mathbf{F}$ is the force, and χ is the response function.

If the system is translationally invariant, or can be approximated as so, then χ is only a function of the difference coordinate $\mathbf{r} - \mathbf{r}'$, rather than either coordinate separately; and if the system is already in equilibrium, then the response can only depend on the time difference $t - t'$. So Eq. (4.1) simplifies to

$$u_\alpha(\mathbf{r},t) = \int d\mathbf{r}' \int dt' \chi_{\alpha\beta}(\mathbf{r}-\mathbf{r}',t-t') F_\beta(\mathbf{r}',t') \ , \tag{4.2}$$

which is a convolution in space and time. The relationship becomes simpler by Fourier transforming in both space and time

$$u_\alpha(\mathbf{q},\omega) = \chi_{\alpha\beta}(\mathbf{q},\omega) F_\beta(\mathbf{q},\omega) \ . \tag{4.3}$$

What is not quite so obvious is that the response function also tells us about *collective modes* in the solid.

We will continue with the formalities in a moment. It is best to get a feeling for the concepts with some examples. The first two of these should already be familiar.

4.1.1 Phonons

We do not necessarily need quantum mechanics to describe collective modes, and a familiar example is that of sound waves described by the classical equation of motion for the displacement u (e.g. for waves along a string)

$$-K\nabla^2 u + \rho\ddot{u} + \gamma\dot{u} = F \tag{4.4}$$

where K is the stiffness, ρ the mass density, and γ a damping parameter. $F(r,t)$ is the external force applied to the medium.

Let us first consider the possible wave solutions of Eq. (4.4) in the absence of an external force:$F = 0$. These solutions are the modes of free oscillation - or collective modes of the system. Formally, we can guess that these will be waves of the form

$$e^{i\mathbf{q}\cdot\mathbf{r}}e^{i\omega t} \tag{4.5}$$

and upon substituting this form into Eq. (4.4) we shall find that such a solution is allowed if there is a relation between the wave-vector $\mathbf{q}$ and the frequency ω, namely[1]

$$Kq^2 - \rho\omega^2 + i\gamma\omega = 0 \quad . \tag{4.6}$$

Such a relation between the frequency and wavevector $\omega(\mathbf{q})$ is known as a dispersion relation; here it gives the propagation frequency (and damping, corresponding to the imaginary part of the frequency) as a function of the sound wavelength $2\pi/k$.

Now consider the driven problem $F \neq 0$. The solution by Fourier transforms is straightforward :

$$u(\mathbf{q},\omega) = \frac{F(\mathbf{q},\omega)}{-\rho\omega^2 + i\gamma\omega + Kq^2} \tag{4.7}$$

so the response function is here

$$\chi(\mathbf{q},\omega) = \frac{1}{-\rho\omega^2 + i\gamma\omega + Kq^2} \tag{4.8}$$

which has poles at $\omega = \pm\sqrt{K/\rho}\,q$ in the absence of damping. When damping is present, the poles move into the lower half plane (as they must, due to causality, see below), but provided γ is small the decay rate will be slow. Notice that the poles of the response function occur exactly along the dispersion curve Eq. (4.6).

Remember that the equation of motion Eq. (4.4) can be derived from a classical Hamiltonian (in the absence of dissipation), which is

$$H = \int d\mathbf{r}\left[\frac{1}{2}K(\nabla u)^2 + \frac{1}{2}\rho(\dot{u})^2 + Fu\right] \quad , \tag{4.9}$$

---

[1] See the appendix for a reminder

where we have included the external force F as a perturbation. A special feature of this Hamiltonian is that there is no term directly proportional to the displacement – only to its derivatives in space and time. Physically this is because of a symmetry – the solid as a whole is translationally invariant and the energy cannot depend on arbitrary uniform displacements. We could have such a term if u were not the coordinate of a sound wave, but say of an optical phonon, when there would be an additional term in the Hamiltonian proportional $\frac{1}{2}Vu^2$.

For acoustic phonons, the absence (by symmetry) of such terms means that $\omega_o(q) \propto q$, whereas for an optic mode $\lim_{q\to 0}\omega_o(q) \to$ const.. We sometimes say that acoustic modes are massless, and others massive.

Phonon spectra can be mapped out by inelastic scattering probes, and most efficient is neutron scattering, because the characteristic wavelength of the neutrons can be easily made comparable to the lattice constant of the solid. So there is relatively efficient scattering from phonon modes of all wavelengths. Inelastic light scattering from light in the visible or UV range (Raman scattering) can only see optical modes near $q = 0$, because the wavelength of visible light is very long. However, the advent of bright synchrotron X-ray sources means that inelastic X-ray scattering will be an important probe (for this, and other modes) in the future.

4.1.2 Plasmons

Another classical oscillation that follows the same prescription is the plasma oscillation of the electron gas. A plasma oscillation is a longitudinal density wave, with the electrons displacement of the form

$$\mathbf{u} = \mathbf{u}_o e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)} \quad \mathbf{u}_o \parallel \mathbf{q} \tag{4.10}$$

where the longitudinal character has been chosen by placing the displacement parallel to the wavevector. The longitudinal displacement of the charge induces a polarisation

$$\mathbf{P} = -ne\mathbf{u} \tag{4.11}$$

and therefore an internal electric field in the sample which is

$$\mathbf{E} = -4\pi\mathbf{P} + \mathbf{D} \tag{4.12}$$

where $\mathbf{D}$ is the displacement field, due to external forces only. $\mathbf{D}$ is the applied field. The equation of motion for the displacement is

$$m\ddot{\mathbf{u}} = -e\mathbf{E} = -4\pi ne^2\mathbf{u} - e\mathbf{D} \tag{4.13}$$

which after rearrangement and Fourier transformation gives

$$(-m\omega^2 + 4\pi ne^2)u = -eD \ . \tag{4.14}$$

The response function – here the inverse of the factor on the left hand side of Eq. (4.14) has a pole at the plasma frequency $\omega_p = (4\pi ne^2/m)^{1/2}$ – which describes a massive mode[2].

It is more conventional to use a response function to relate the internal electric field **E** to the applied field **D**: some straightforward reshuffling leads to

$$\left[1 - \frac{\omega_p^2}{\omega^2}\right] E_\parallel = D_\parallel \ , \tag{4.15}$$

which defines the longitudinal dielectric function

$$\epsilon_\parallel = \left[1 - \frac{\omega_p^2}{\omega^2}\right] \ . \tag{4.16}$$

Notice that the response function relating the internal field to the external field is however actually $1/\epsilon$. Since D is generated (by Poisson's law) by the external potential, and E is generated by the screened potential, another way of defining the dielectric function is

$$V_{sc}(\mathbf{q},\omega) = \frac{V_{ext}(\mathbf{q},\omega)}{\epsilon(\mathbf{q},\omega)} \tag{4.17}$$

Defined this way, we have already estimated a formula for the static dielectric function in Eq. (2.146).

This classical discussion does not generate any dispersion for the plasmon, i.s. the plasma frequency is found to be q-independent. It turns out that the classical theory is *exact* for $\mathbf{q} \to 0$, but there are quantum excitations that are entirely missed at short wavevectors.

Since ϵ measures the charge response of a solid, then plasmons are generated by any charged probe. The classic experiment to observe plasmons is *Electron Energy Loss Spectroscopy* (EELS), where a high energy electron is sent into the sample and the energy loss monitored; an EELS spectrum will therefore be proportional to

$$\Im \frac{1}{\epsilon(\mathbf{q} \approx 0, \omega)} \tag{4.18}$$

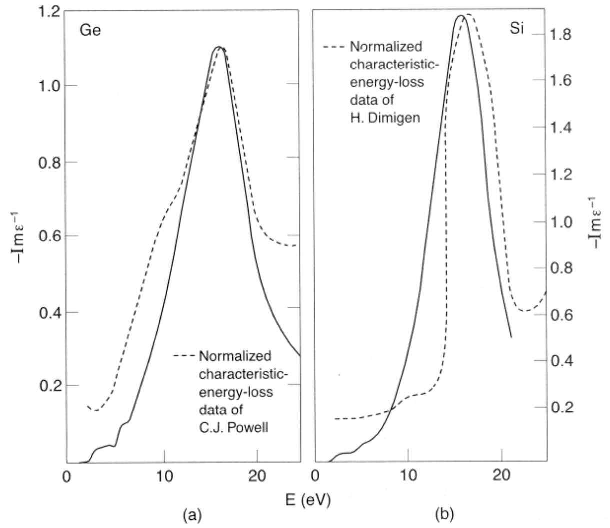
which will therefore have a peak near the plasma frequency.

4.1.3 Optical conductivity of metals

The equation of motion we wrote down for the electrons in a solid Eq. (4.14) assumes that electrons are accelerated in an applied field, and do not suffer any damping — it would predict the conductivity of a metal to be infinite. This is fixed up in a *Drude model* by adding a phenomenological damping

---

[2] In SI units $\omega_p^2 = ne^2/\epsilon_o m$



(a) (b)

Figure 4.1: . Electron energy loss spectrum for Ge and Si (dashed lines) compared to values of $Im(1/\epsilon)$ extracted directly from measurements of the optical conductivity. [From H.R.Philipp and H.Ehrenreich, Physical Review **129**, 1550 (1963)

term, that represents the effects of scattering. With the additional term, the modified equation of motion becomes

$$m\ddot{\mathbf{u}} + \gamma\dot{\mathbf{u}} = -e\mathbf{E} = -4\pi ne^2\mathbf{u} - e\mathbf{D} \tag{4.19}$$

which after Fourier transformation gives

$$(-m\omega^2 - i\gamma\omega + 4\pi ne^2)u = -eD \quad . \tag{4.20}$$

Clearly we can rework the previous analysis by replacing everywhere $m\omega^2$ by $m\omega^2 + i\gamma\omega$. We will then arrive at a *complex* dielectric function

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau} \quad , \tag{4.21}$$

where we have defined a relaxation rate

$$\frac{1}{\tau} = \frac{\gamma}{m} \quad . \tag{4.22}$$

This is one way of expressing the result. An alternative expression is not to relate the displacement field to the electric field, but instead to calculate the current

$$j = -ne\dot{u} = i\omega neu \tag{4.23}$$

induced by the applied electric field. After a few algebraic manipulations, we get

$$j(\omega) = \frac{\omega_p^2}{4\pi(1/\tau - i\omega)} E(\omega) \quad . \tag{4.24}$$

This instead expresses the result as a complex conductivity

$$\sigma(\omega) = \frac{\omega_p^2}{4\pi(1/\tau - i\omega)} \quad . \tag{4.25}$$

One must realise that the two results Eq. (4.21) and Eq. (4.25) are completely equivalent: a relation between D and E implies a relation between j and E. One may translate between the two formulae by[3]

$$\epsilon(\omega) = 1 + i\frac{4\pi\sigma(\omega)}{\omega} \quad . \tag{4.26}$$

The real part of the conductivity is proportional to the imaginary part of the dielectric function.

From Eq. (4.25), one can see that the theory gives rise to a familiar d.c. conductivity

$$\sigma(0) = \frac{\omega_p^2\tau}{4\pi} = \frac{ne^2\tau}{m} \quad , \tag{4.27}$$

so that τ has a simple interpretation as the mean free time between collisions. At frequencies larger than $1/\tau$ the conductivity rapidly falls off:

$$\Re\sigma(\omega) = \frac{\sigma(0)}{1+\omega^2\tau^2} \quad . \tag{4.28}$$

Consequently, the dielectric function near $q = 0$ can be extracted directly from optical absorption of light, which gives a measurement of the ac conductivity $\sigma(\omega)$. A comparison between optical measurements and electron energy loss measurements is shown in Fig. 4.1.

4.1.4 Static response

The familiar static susceptibilities in magnetism and dielectrics are just the response to a field that is uniform in both time and space – viz $\chi(\mathbf{q} \to 0, \omega \to 0)$. In a classical thermodynamic system the response function can always be determined by writing down the Helmholz free energy in the presence of an external field: we shall largely be concerned with zero temperature, so we need only to consider the internal energy

$$\mathcal{U} = U_o(\mathbf{u}) - \mathbf{u}\cdot\mathbf{F} \quad , \tag{4.29}$$

where U_o is the energy in the absence of the external field, and the last term represents the work done by the external field on the system. We shall be

---

[3] In SI units, replace the 4π by $1/\epsilon_o$

usually dealing with the case when the response vanishes in the absence of the external field. Furthermore, we are interested in the *linear* response, so we shall assume that all the fields are small, so that U_o can be expanded in a Taylor series (simplifying to a scalar field, but the generalisation is straightforward):

$$U_o(u) = U_o + U_o'\, u + \frac{1}{2} U_o''\, u^2 + ... \tag{4.30}$$

Minimising the energy with and without a field leads then to the stability condition in the absence of the field

$$U_o' = 0 \quad , \tag{4.31}$$

and the static susceptibility

$$u = \frac{1}{U_o''} F \quad . \tag{4.32}$$

This classical prescription for calculating a response function can be directly adapted to quantum mechanics, and in fact quantum mechanics offers a prescription for calculating the response to time-varying fields that is much more straightforward than in the classical thermodynamic case.

One last remark, that we will return to later: since U_o'' depends on many parameters which could perhaps be smoothly changed – e.g. pressure, temperature – it might be possible that this could vanish and change sign. In that case, we chould see $\chi = 1/U_o''$ diverge, and at the sign change in U_o'' the displacement will grow – not without bound, because eventually higher-order terms in Eq. (4.30) will become relevant. This is a continuous (or second order) phase transition, where the susceptibility diverges at the critical point. A very important example of this is discussed in a question - the Landau theory of phase transitions.

4.1.5 Dynamic response, collective modes, and particles

The examples above should have given us a lot of intuition about the relationship between response functions and collective modes. If there exists a curve $\omega = \omega_o(\mathbf{q})$ where $\chi(\omega, \mathbf{q})$ is divergent, then if a driving force is applied at exactly the correct frequency and wavevector to match this dispersion curve, the response u will be infinite; another way of putting this is to say that even for an infinitesimal driving force there will be a response. The dispersion curve $\omega = \omega_o(\mathbf{q})$ defines modes of *free oscillation* of the system.

These collective modes may be directly visible in an inelastic scattering experiment (see Fig. 4.2) If one sends in an incident probe that has the appropriate symmetry to couple to the collective excitations, it will then be able to create (or absorb) the collective modes, and therefore be scattered

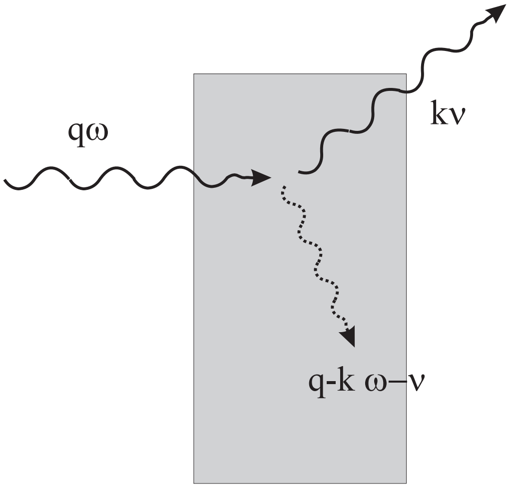


Figure 4.2: Generic description of an inelastic scattering experiment

both in frequency and energy. By comparing the incident and outgoing energy and momenta, one can then deduce the spectrum of excitations. If the excitations are underdamped (poles close to the real frequency axis) then the features will appear as sharp resonances; continuum excitations will lead to broad features. Very generally , the probability of making an excitation of momentum $\mathbf{q}$ and energy ω is proportional to

$$\Im\chi(\mathbf{q},\omega) \tag{4.33}$$

of course multiplied by matrix elements and selection rules appropriate for the probe at hand. [4]

The collective modes are just waves propagating through the solid; in the examples above they are waves of atomic vibration (phonons), of longitudinal polarisation (plasmons),or of transverse electric field (light). Later on we shall see other examples, e.g. spin waves.

4.1.6 Causality and Kramers-Krönig relations

The response functions we are using must be *causal* or *retarded*, so that there is no possibility of a response before the force is applied, i.e. a causal response function κ must satisfy

$$\kappa(t-t') = 0 \quad \text{if} \;\; t' > t \;\; . \tag{4.34}$$

The principle of causality imposes conditions on the behaviour of $\kappa(\omega)$ in Fourier space (we shall drop the momentum or space coordinate for the moment): $\kappa(\omega)$ must be an *analytic* function of ω in the upper half plane,

---

[4] We shall use the symbols $\Im$ (Imaginary part), $\Re$ (Real part) and $\wp$ (Principal value).

and in turn this analyticity will enforce a relation between the real and imaginary parts of $\kappa(\omega)$ on the real frequency axis, which are called Kramers-Krönig relations.

Derivation of Kramers-Krönig relations.

You should understand the principles behind this, but do not need to be able to reproduce the derivation. For those who need reminding, some elements of the theory of analytic functions are in the appendix.

$\kappa(t)$ and its Fourier transform are related by the standard relations

$$\kappa(\omega) = \int_{-\infty}^{\infty} \kappa(t) e^{i\omega t} dt \qquad \kappa(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \kappa(\omega) e^{-i\omega t} d\omega \quad . \tag{4.35}$$

We may construct $\kappa(t)$ by evaluating the inverse Fourier transform (Eq. (4.35) as a contour integral, as shown in the Fig. 4.3. Provided $|\kappa(\omega)|$ falls off faster than $1/\omega$ as $|\omega| \to \infty$, then we can close the contour by a semicircle at ∞, in either the upper or lower half plane, depending as $t < 0$ or $t > 0$ respectively.

Then the contour can be shrunk to enclose the poles, so that

$$\begin{aligned}\kappa(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \kappa(\omega) e^{-i\omega t} d\omega \\ &= 2\pi i \sum (residues)\end{aligned} \tag{4.36}$$

In order for causality, and Eq. (4.34) to be satisfied, $\kappa(\omega)$ must be an analytic function in the upper half plane, so that there are no poles contributing to the integral. All the non-analytic structure of the response function must lie below the real frequency axis.

Using this property, we can write (using Cauchy's theorem)

$$\kappa(\omega) = \oint \frac{d\omega'}{2\pi i} \frac{\kappa(\omega')}{\omega' - \omega - i\eta} \tag{4.37}$$

for any causal response function $\kappa(\omega)$, where the contour runs as shown in Fig. 4.4. We should remember that $\kappa(\omega)$ is a complex function, even on the real frequency axis, and it is useful to use Eq. (4.37) to derive an integral relationship between the real and imaginary parts. To do this, we need to have all the terms multiplying $\kappa(\omega')$ to be pure imaginary. We thus send $\eta \to 0$, deforming the contour as shown in Fig. 4.4. If we perform the integral around the semicircle it gives precisely *half* the contribution of the pole, and what remains is the *principal value* integral along the line:

$$\kappa(\omega) = \wp \int_{-\infty}^{\infty} \frac{d\omega'}{\pi i} \frac{\kappa(\omega')}{\omega' - \omega} \tag{4.38}$$

Taking the real and imaginary parts separately of Eq. (4.38) we get the *Kramers-Krönig relations*:

$$\begin{aligned}\Re\kappa(\omega) &= \wp \int \frac{d\omega'}{\pi} \frac{\Im\kappa(\omega')}{\omega' - \omega} \\ \Im\kappa(\omega) &= -\wp \int \frac{d\omega'}{\pi} \frac{\Re\kappa(\omega')}{\omega' - \omega}\end{aligned} \tag{4.39}$$

The real and imaginary parts are not independent.

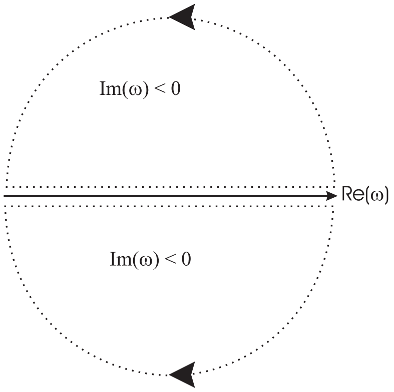


Figure 4.3: Contour to evaluate the inverse Fourier transform of $\kappa(\omega)$. For $t < 0$, the contour is closed in the upper half plane; for $t > 0$ in the lower half plane

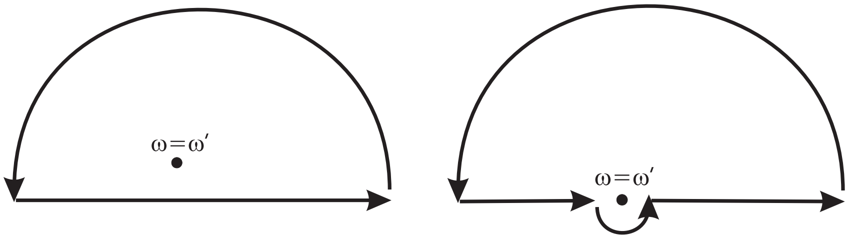


Figure 4.4: Contour to evaluate the Cauchy integral Eq. (4.37). Upon deforming the contour so that the pole moves onto the real axis, the integral separates into a principal value integral along the real line (defined by integrating along the domain $(-\infty, \omega - \eta][\omega + \eta, \infty))$, and the contribution integrating half way around the pole, leading to the difference of a factor of two between Eq. (4.37) and Eq. (A.75)

The KK-relations are not just a mathematical nicety, but a practically useful tool. For example, one can infer the frequency-dependent complex conductivity simply by measuring the reflectivity as a function of frequency (all phase information about the complex reflection coefficients is apparently lost), and then reconstruct the full complex conductivity using Eq. (4.39).

4.1.7 Recap.

It is worth listing the general properties of dynamical response functions that we have discussed in this section.

- **Definition.** A linear response function $\chi(\mathbf{q}\omega)$ in a homogeneous medium defines the response u to an infinitesimal applied force F.

$$u(\mathbf{q}, \omega) = \chi(\mathbf{q}, \omega) F(\mathbf{q}, \omega) \quad . \tag{4.40}$$

- **Collective modes.** Modes of free oscillation of the system are described by the pole structure of χ; namely the dispersion $\omega(q)$ of the collective mode is given by the solutions of

  $$\chi(\mathbf{q}, \omega(\mathbf{q}))^{-1} = 0 \quad . \tag{4.41}$$

  The solution for $\omega(\mathbf{q})$ may be complex, in which case the modes are damped.

- **Instabilities.** If the static susceptibility diverges (e.g. as a function of some parameter), it is a signature of an instability of the system.

- **Energy absorption.** The rate of energy dissipated by the system when driven by an external force $F \exp(i\mathbf{q}\cdot\mathbf{r} + i\omega t)$ is proportional to $\Im\chi(\mathbf{q}, \omega)$.

- **Causality.** A causal response function $\chi(\omega)$ is an analytic function of ω in the upper half frequency plane. The real and imaginary parts of χ are related by integral relations called the Kramers-Krönig relations, and therefore are not independent.

4.2 The electron as a collective excitation

The title of this section seems rather perverse, because we surely know that an electron is a fundamental particle. But in a solid, the an *individual* electron cannot be separated from the rest of the interacting quantum liquid, so the existence of individual particles is moot. Instead, we should think of the kind of experiments we might do to study "electron-like" excitations.

The most direct thing we could do is to add an electron at some point $\mathbf{r}$ at time t, and take an electron out at some later time t' at a point $\mathbf{r}'$. What we would be studying is the propagation of an electron from one point to another in a solid. As always, it is easier to think of this in Fourier ($\mathbf{k}\omega$) space, and let us start with free particles with no interaction between them.

If we put a particle into an eigenstate of the Hamiltonian labelled by its momentum $\mathbf{k}$, then the wavefunction will evolve in time following the Schrödinger prescription

$$\psi_k(\mathbf{r}, t) = \psi_k(\mathbf{r}) e^{-i\epsilon_{\mathbf{k}} t} \quad . \tag{4.42}$$

Here ψ_k is the Bloch wavefunction satisfying the time-independent Schrödinger equation, and the time-dependent solution oscillates in time with a single frequency $\omega = \epsilon_{\mathbf{k}}$. If we look at this in fourier space, we should say that

$$\psi_k(\mathbf{r}, \omega) = 2\pi\psi_k(\mathbf{r})\delta(\omega - \epsilon_{\mathbf{k}}) \tag{4.43}$$

so that the wavefunction has spectral weight only at $\omega = \epsilon_{\mathbf{k}}/\hbar$. [5]

We can say that the probability of finding an electronic state with energy ω and momentum $\mathbf{k}$ is just

$$A(\mathbf{k},\omega) = \delta(\omega - \epsilon_{\mathbf{k}}) \quad , \tag{4.44}$$

and the quantity $A(\mathbf{k},\omega)$ is usually called the *electron spectral function.*

What about an interacting system? If we add a particle, it will collide, interact, and exchange with all the other particles in the system, but supposing the interactions are weak (though they aren't) one might imagine that something similar to the particle will survive. First, if we fix the momentum $\mathbf{k}$ of the excitation, that cannot change because the two-particle interactions conserve momentum. Of course, some of the character of the original particle will now be shared with a "cloud" of electrons screening it. Second, since we can see that in a metal there is a continuum of excitations, we should expect it to decay with a finite lifetime - say $1/\Gamma_{\mathbf{k}}$. The time-dependence of the state will now become $e^{(i\epsilon_{\mathbf{k}}-\Gamma_k)t}$. Third, there is no reason that the dispersion relation should be the same as for the free particle: we should replace $\epsilon_{\mathbf{k}}$ by a renormalised $\tilde{\epsilon_{\mathbf{k}}}$. Last, one might imagine that because there are already electrons in the system, it is in some sense harder to find space for it, and its overall spectral weight might be reduced. These effects can be combined into an *ansatz* for the spectral function in an interacting system of the form

$$A(\mathbf{k},\omega) = -\frac{1}{\pi}\Im\left[\frac{Z_{\mathbf{k}}}{\omega - \tilde{\epsilon_{\mathbf{k}}} + i\Gamma(\mathbf{k})}\right] \tag{4.45}$$

Notice that if the inverse lifetime $\Gamma_{\mathbf{k}} \to 0$ and the spectral weight $Z_{\mathbf{k}} \to 1$, the spectral function reduces to the noninteracting Eq. (4.44). Eq. (4.45) describes *quasiparticles* with spectral weight $Z_{\mathbf{k}} \leq 1$, a dispersion curve $\tilde{\epsilon_{\mathbf{k}}}$, and a decay rate (inverse lifetime) $\Gamma_{\mathbf{k}}$. If $\Gamma_{\mathbf{k}}$ is small, then the quasiparticles are long-lived and have some real meaning. They also have charge e and spin $\frac{1}{2}$; we will not mess with Fermi statistics.

We must be careful about the chemical potential. If we are in equilibrium (and at T=0), we cannot add fermionic excitation at an energy $\omega < \mu$. So we shall infer that for $\omega > \mu$, Eq. (4.45) is the spectral function for particle-like excitations, whereas for $\omega < \mu$ it is the spectral function for *holes.*

4.2.1 The Fermi liquid

Why is it that Γ can be small in a metal where the typical separation between electrons is only an Angstrom or so? The answer is provided by Fermi statistics, and is codified in Landau's theory of the Fermi liquid. In

---

[5] Getting the sign here requires one to adopt a sign convention for Fourier transforms that is used in Eq. (4.35) – opposite to the one often used in maths books. We have also set $\hbar = 1$.

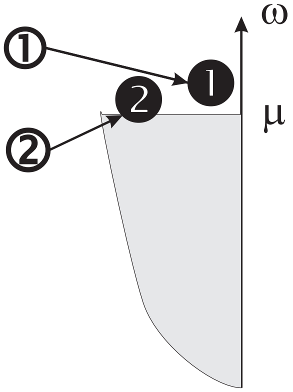


Figure 4.5: Collision between two quasiparticles near the Fermi surface. The initial and final states are marked by open and filled circles.

short, the fact that Fermi statistics exclude double occupancy of the same quasiparticle state *guarantees* that

$$\Gamma \propto (\omega - \mu)^2 \quad . \tag{4.46}$$

So for excitations close to the fermi surface, the lifetime becomes very long.

Here is a sketch of how to obtain that result. Let us assume that quasiparticles exist, and estimate the effect of the interactions between them. So in Fig. 4.5 we show the Fermi sea, with a test particle (1) added – of course it has to be placed above the chemical potential because the states below are filled. Now consider the interaction of this particle with a particle (2) (which must be inside the Fermi sea). The result of the collision will be scattering to final states labelled ($1'$, $2'$). The final states must have initially been empty (Fermi statistics), so both $1'$ and $2'$ must lie at energies above the chemical potential μ. But we also have to satisfy energy (and momentum) conservation

$$\omega_1 + \omega_2 = \omega_{1'} + \omega_{2'} \quad , \tag{4.47}$$

which means that

$$\omega_1 + \omega_2 > 2\mu \quad . \tag{4.48}$$

We can rewrite this as

$$\mu - \omega_2 < \omega_1 - \mu \quad . \tag{4.49}$$

We see that the only particles which are allowed to scatter from the test particle 1 are those whose energy is closer to the Fermi energy than particle 1 itself. So if $\omega_1 \rightarrow \mu$, there are no scattering processes allowed, and the quasiparticle is long-lived.

There is a further constraint according to momentum conservation. The momentum transfer in the scattering process takes particle 1 from state $\mathbf{p}$ to $\mathbf{p} + \mathbf{\Delta p}$, with exactly the same momentum transfer in reverse for particle 2. We can separate the components of the momentum transfer perpendicular and parallel to the Fermi surface, and it is clear that $\mathbf{\Delta p}_\perp < v_F(\omega_1 - \mu)$, where v_F is the Fermi velocity. In order to calculate the overall scattering rate, we need to integrate over all the possible final states, and over all the possible states of particle 2. What is important in what follows is that two of these integrals are constrained: the energy of particle 2, and the momentum transfer perpendicular to the FS. We can now estimate that the scattering rate must be of order

$$\Gamma(\omega_1) \quad \propto \quad \int_{2\mu-\omega_1}^{\mu} d\omega_2 \int_0^{v_F(\omega_1-\mu)} d\mathbf{\Delta p}_\perp \int d(other\ momenta) W(1,2;1',2')$$

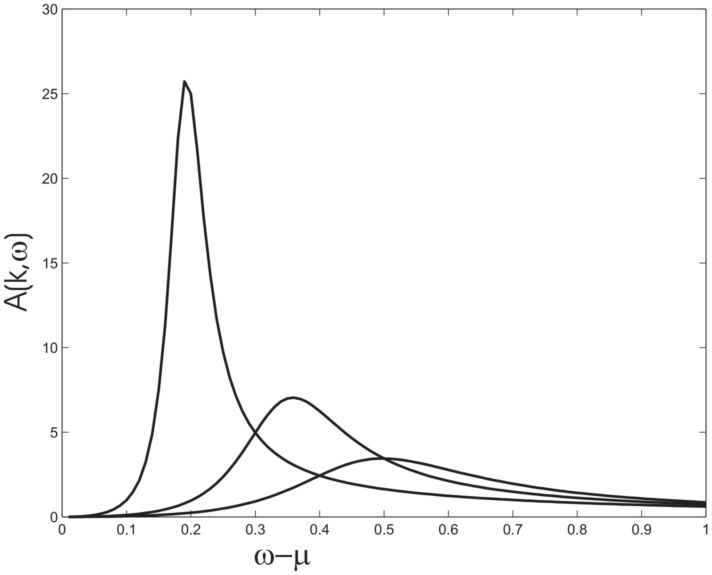


Figure 4.6: Expected form of the quasiparticle spectral function $A(\mathbf{k}, \omega)$ plotted for momenta crossing through the Fermi energy. The spectral function is plotted as a function of energy, for three different momenta corresponding to bare energies $\epsilon_{\mathbf{k}} - \mu = 0.2$, 0.4 and 0.6; the damping rate is assumed to be of the fermi liquid form $\Gamma = \frac{1}{2} W (\omega - \mu)^2$, with here $W = 2$ as an example.

$$\propto \quad W(\omega_1 - \mu)^2 \tag{4.50}$$

where W is the scattering matrix element, which we replace by its average value near the Fermi surface. So $\Gamma \to 0$ for quasiparticles whose energy lies exactly on the Fermi surface, and our assumption that quasiparticles exist and are long-lived is self-consistent. Moreover, it can be shown that the quasiparticles are in a one-to-one correspondence with the particles of the noninteracting theory.[6]

Fig. 4.6 shows a plot of a model quasiparticle spectral function as a finction of momentum and energy passing through the Fermi momentum. There is a peak in the spectral function along the dispersion curve $\omega = \epsilon_{\mathbf{p}}$, which sharpens as the Fermi energy is approached.

This is an existence proof[7] of the quasiparticle concept, but of course it does not guarantee that the Fermi liquid state always exists. It can be shown that turning on the interaction between particles adiabatically from zero then the free particles smoothly transform into quasiparticles,

---

[6] The proof is formidable, and is known as Luttinger's theorem : J.M.Luttinger and J.C.Ward, *Physical Review* **118**, 1417 (1960); *ibid.* **119**, 1153 (1960).

[7] Actually not very rigorous. The rigorous basis is given in the book by A.A.Abrikosov, L.P.Gor'kov, and I.E.Dyalozhinski, *Methods of Quantum Field Theory in Statistical Physics* , Dover Edition, 1975

and the volume contained within the Fermi surface is unchanged (this is the Luttinger theorem). However, this does not preclude a phase transition, where the Fermi liquid character abruptly disappers. The conditions under which fermi liquids exist or not is an active field of both experimental and theoretical research.

4.2.2 Photoemission

The most direct way to measure the electron spectral function directly is by photoemission, although this is a difficult experiment to do with high resolution. In a photoemission experiment, photons are incident on a solid, and cause transitions from occupied states to plane wave-like states well above the vacuum energy; the excited electron leaves the crystal and is collected in a detector that analyses both its energy and momentum.[8] The photon carries very little momentum, so the momentum of the final electron parallel to the surface is the same as the initial state in the solid, while of course the perpendicular component of the momentum is not conserved. Photoemission data is therefore most easy to interpret when there is little dispersion of the electronic bands perpendicular to the surface, as occurs in anisotropic layered materials. It is fortunate that there are many interesting materials (including high-temperature superoconductors) in this class. If one analyses both the energy and the momentum of the outgoing electron, (this is Angle Resolved Photo-Emission Spectroscopy, or ARPES) the signal will be proportional to

$$I_{\mathbf{k}} A(\mathbf{k},\omega) f(\omega) \tag{4.51}$$

where $I_{\mathbf{k}}$ is the (dipole) matrix element for transitions, and is slowly varying as a function of momentum, and $f(\omega)$ is the Fermi function – one can have emission from occupied states only.

If one integrates over all angles, then the spectrum is roughly proportional to the density of states in energy (exactly so if the matrix element is a constant), i.e.

$$\sum_{\mathbf{k}} A(\mathbf{k},\omega) f(\omega) = N(\omega) f(\omega) \quad , \tag{4.52}$$

measuring the energy-dependent density of states $N(\omega)$. The ideal schematic for an ARPES experiment would then be as shown in Fig. 4.6, cutoff by the fermi function so that the line should disappear as the Fermi surface is crossed.

An example of real data is shown in Fig. 4.9.

---

[8] For a detailed discussion of photoemission experiments, see Z.X.Shen and D.S.Dessau, *Physics Reports*, **253**, 1-162 (1995)

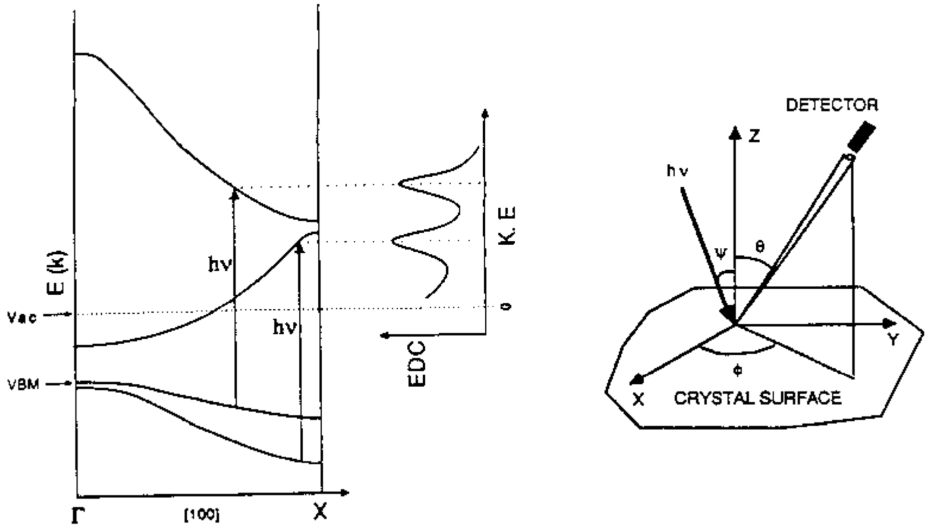


Figure 4.7: Schematics of a photoemission experiment. The optical transitions are nearly vertical, so the electrons are excited from the valence bands to high energy excited states (above the vacuum energy necessary to escape from the crystal) with the same crystal momentum. In this case the two transitions that can be excited by a single frequency will yield a double peak in the kinetic energy distribution of the escaped electrons. When the excited electrons escape through the surface of the crystal, their momentum perpendicular to the surface will be changed. If the surface is smooth enough, the momentum of the electron parallel to the surface is conserved, so the angle of the detector can be used to scan $k_{\|}$

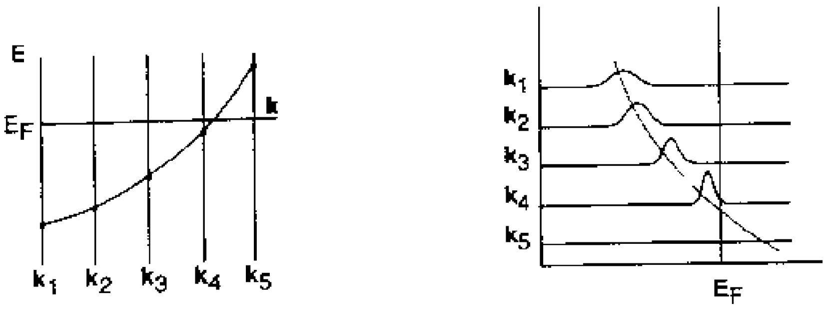


Figure 4.8: Idealised results from a photoemission experiment. A peak is observed at the band energy in each spectrum, but disappears when the band crosses the Fermi energy

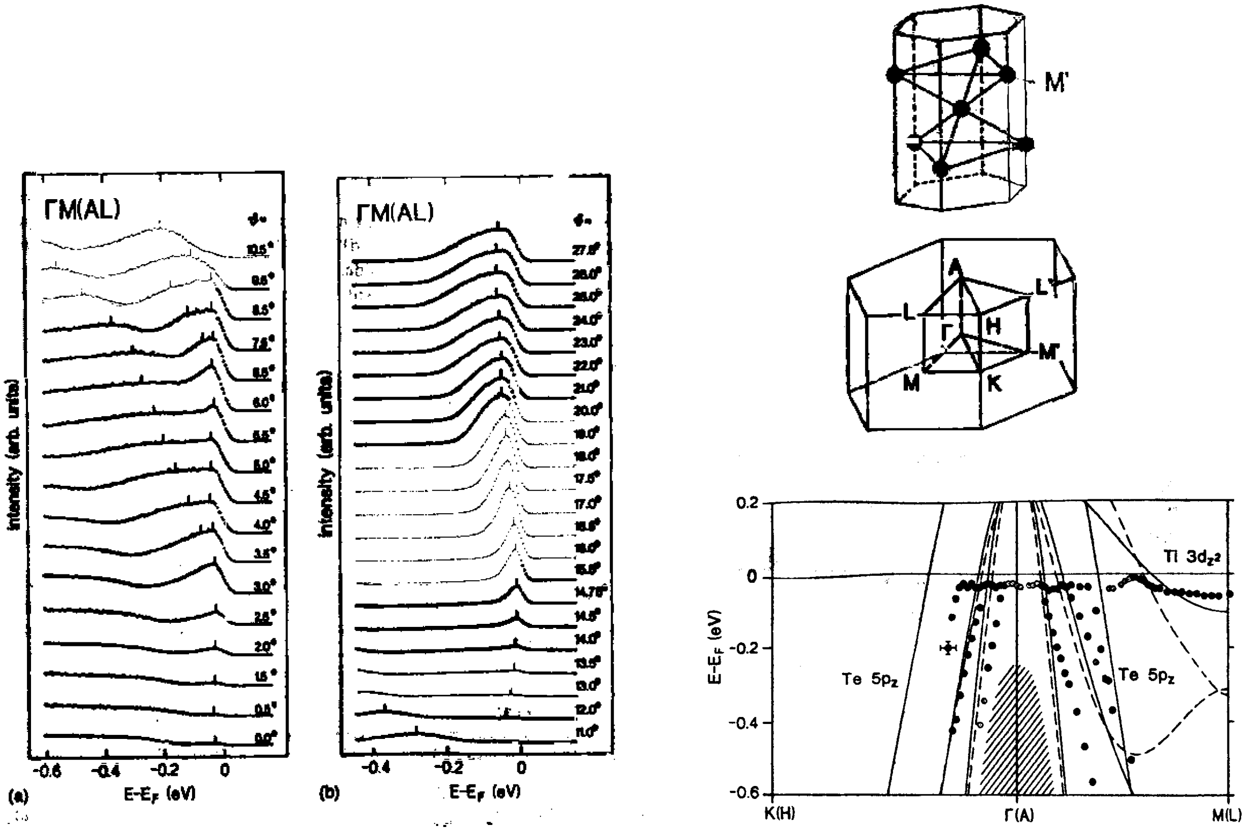


Figure 4.9: . Photoemission spectra on the two dimensional layered metal $TiTe_2$. The bands are nearly two-dimensional in character, so the interpretation of the photoemission data is straightforward – different angles (see Fig. 4.7)correspond to different in-plane momenta. The left panels show energy scans as a function of angle that correspond to changing the in-plane momentum in the direction from the centre of the Brillouin zone Γ towards the centre of the hexagonal zone face. (The right hand upper panels show the crystal unit cell, and the reciprocal lattice unit cell.) Several bands appear to cross the Fermi energy, with very different velocities, and there is some tendency to sharpening of the features as their energies approach E_F. The right hand lower panel plots the positions of the peaks as a function of energy and momentum in comparison to the band structure (on a very expanded scale). Open circles correspond to weak features in the data, and may not be real. Steeply rising bands correspond to rapidly dispersing states derived mostly from Te $5p$ orbitals, and a nearly flat band appears to arise from the Ti $3d$ states. From Claessen et al., *Physical Review B*, **54**, 2453, (1996).

4.3 Dynamics of the electron gas

4.3.1 Density response function

We now need to ask what the response of the electron gas is to a time dependent potential, so that we can calculate the density response function, viz.

$$\delta\rho(\mathbf{q},\omega) = \chi(\mathbf{q}\omega)V(\mathbf{q},\omega) \quad , \tag{4.53}$$

We will start by doing this for free electrons, but including Fermi statistics; following this we will put in the interaction effects only via the screening of the potential, in a way that should be familiar already from the discussion of screening earlier.

The next three sections sketch the derivation of the density response function for the free Fermi gas. They are all straightforward (but messy) algebra, and you are not expected to be able to reproduce them. The end results are transparent, but you may find it worth going through the equations at least once.[9]

Time-dependent perturbation theory

Our calculations will use standard time-dependent perturbation theory; this is completely straightforward, as we need to go only to first order, with the main problem being to keep track of all the indices. Here is a quick reminder of the standard results from quantum mechanics.

We have to solve the Schrodinger equation under a time-dependent perturbation $V(\mathbf{r},t)$ which is assumed small, i.e.

$$i\hbar\frac{\partial\psi(\mathbf{r},t)}{\partial t} = \left[H_o + V(\mathbf{r},t)\right]\psi(\mathbf{r},t) \ . \tag{4.54}$$

Assume that we have already solved for the eigenvalues ϵ_n and eigenfunctions $\phi_n(\mathbf{r})$ of H_o. Thus we now look for a solution of the time-dependent problem in the form

$$\psi(\mathbf{r},t) = \sum_n a_n(t)e^{-i\epsilon_n t/\hbar}\phi_n(\mathbf{r}) \tag{4.55}$$

where we now need to find the time-dependent coefficients a_n. As always, substituting back in the SE, and taking matrix elements w.r.t. the basis functions gives us the following

$$i\hbar\frac{\partial a_m(t)}{\partial t} = \sum_n a_n(t)e^{i\omega_{mn}t} < m|V|n > \quad . \tag{4.56}$$

where $\hbar\omega_{mn} = \epsilon_m - \epsilon_n$. There are no approximations so far, but now the perturbation theory arrives. We shall assume that the system starts out at $t = 0$ in a particular state $a_n(t=0) = a_n^0 = 1$; then we get to first order in V

$$a_m^1(t) = -\frac{i}{\hbar}\int dt < m|V|n > e^{i\omega_{mn}t}a_n^{(0)} \quad . \tag{4.57}$$

---

[9] For more details of the general method, a good book is *Landau and Lifshitz*, Statistical Physics, A course of Theoretical Physics, vol 5, pages 377-388.

Now let us choose a particular form of the perturbation to be just a plane wave

$$V(\mathbf{r},t) = V e^{i\mathbf{q}\cdot\mathbf{r}} e^{-i\omega t} \tag{4.58}$$

and choose as our basis states the free particle plane waves

$$|\mathbf{k}>= e^{i\mathbf{k}\cdot\mathbf{r}} \tag{4.59}$$

So after a little time the first order perturbation theory result is

$$\psi_k(\mathbf{r},t) = |\mathbf{k}> e^{-i\epsilon_{\mathbf{k}}t/\hbar} + a^{(1)}_{\mathbf{k}+\mathbf{q}}(t)|\mathbf{k}+\mathbf{q}> e^{-i\epsilon_{\mathbf{k}+\mathbf{q}}t/\hbar} \tag{4.60}$$

where

$$a^{(1)}_{\mathbf{k}+\mathbf{q}}(t) = \frac{V}{\hbar}\frac{e^{i(\omega_{mn}-\omega)t}}{\omega_{mn}-\omega} \quad . \tag{4.61}$$

Notice that the k in ψ_k is just a label to remind you of where the state originated; because of the scattering, k and $k+q$ are now inextricably mixed.

Now we need to calculate the density change, which is

$$\delta\rho(\mathbf{r},t) = \sum_k \left[|\psi_k(\mathbf{r},t)|^2 - 1\right] \tag{4.62}$$

which is, to linear order in perturbation theory

$$\begin{aligned}
\delta\rho(\mathbf{r},t) &= \sum_k \left[a^{(1)}_{\mathbf{k}+\mathbf{q}}(t) e^{i\mathbf{q}\cdot\mathbf{r}} e^{-i(\epsilon_{\mathbf{k}}-\epsilon_{\mathbf{k}+\mathbf{q}})t/\hbar} + complex\ \ conjugate\right] \\
&= V\sum_k \frac{e^{i\mathbf{q}\cdot\mathbf{r}} e^{-i\omega t}}{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \hbar\omega} + complex\ \ conjugate
\end{aligned} \tag{4.63}$$

Notice that the response is real, of course - the potential at $\mathbf{q},\omega$ generates an inphase response (at $\mathbf{q},\omega$) and an out-of-phase response (at $-\mathbf{q},-\omega$). Hence it is conventional (in this case) to define the response function as the response to a real field; so we add to the potential its complex conjugate [10]

$$V^* = V e^{-i\mathbf{q}\cdot\mathbf{r}} e^{i\omega t} \quad , \tag{4.64}$$

which does not generate any extra harmonics. Substituting for $a^{(1)}_k(t)$ and collecting all the terms, we have

$$\begin{aligned}
\delta\rho &= \sum_k \left[\frac{1}{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \hbar\omega} + \frac{1}{\epsilon_{\mathbf{k}} - \epsilon_{k-q} + \hbar\omega}\right] \times e^{i\mathbf{q}\cdot\mathbf{r}} e^{i\omega t} \\
&+ \ complex\ \ conjugate \quad .
\end{aligned} \tag{4.65}$$

We are almost done. Now we have to account for the occupation of the states, which means that the sum over k should only include states which are occupied in the absence of the perturbation: this means that we can multiply the numerator by a factor $n(k) = 1$ $(|k| < k_F)$ and zero otherwise[11]. Making this replacement, and also writing $\mathbf{k}$ for $\mathbf{k}-\mathbf{q}$ in the second term, we now get[12]

$$\begin{aligned}
\delta\rho &= 2\sum_k \left[\frac{n(\mathbf{k}) - n(\mathbf{k}+\mathbf{q})}{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \hbar\omega}\right] \\
&\times \ e^{i\mathbf{q}\cdot\mathbf{r}} e^{-i\omega t} + complexconjugate
\end{aligned} \tag{4.66}$$

---

[10] This gets rid of the ambiguities about ±ω in the fourier transform

[11] at finite temperature, it is the correct answer to use the Fermi function $n(k) = (e^{\beta(\epsilon_{\mathbf{k}}-\mu)} + 1)^{-1}$.

[12] The factor of 2 is for spin

and hence the density response function is

$$\chi_o(\mathbf{q},\omega) = \frac{\delta\rho(\mathbf{q},\omega)}{V(\mathbf{q},\omega)} = 2\sum_k \frac{n(\mathbf{k}) - n(\mathbf{k}+\mathbf{q})}{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \hbar\omega} \quad . \tag{4.67}$$

Eq. (4.67) is the basic result we shall be using later, but now we will look at another way of getting it.

Energy loss

The time-dependent perturbation theory we started out using can be continued in general to give the famous Fermi's golden rule for transition rates:

$$W_k = \frac{\pi}{\hbar}\sum_l |<k|V(r,t)|l>|^2 \tag{4.68}$$

for the total transition rate out of the state k into all the others (labelled by l) If we apply this formula directly in our case, (i.e. using a plane wave basis) then

$$<\mathbf{k}'|V|\mathbf{k}>= V\left[\delta(\mathbf{k}'-\mathbf{q}-\mathbf{k})\delta(\epsilon_{\mathbf{k}'}-\epsilon_{\mathbf{k}}-\hbar\omega) + \delta(\mathbf{k}'+\mathbf{q}-\mathbf{k})\delta(\epsilon_{\mathbf{k}'}-\epsilon_{\mathbf{k}}+\hbar\omega)\right] \tag{4.69}$$

so that the first δ-function gives conservation of momentum, and the second conservation of energy[13].

In each transition, the system absorbs or emits a quantum of energy $\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{k}} = \pm\hbar\omega$, and the sum

$$Q = V^2\frac{\pi}{\hbar}\hbar\omega\left[\delta(\epsilon_{\mathbf{k}'}-\epsilon_{\mathbf{k}}-\hbar\omega) - \delta(\epsilon_{\mathbf{k}'}-\epsilon_{\mathbf{k}}+\hbar\omega)\right] \tag{4.70}$$

is the mean energy absorbed per unit time. Notice the sign change of the second term relative to the first – absorption versus emission.

If we then sum over all the occupied states $\mathbf{k}$, we have for the total rate of energy dissipation for a perturbation of wavelength $\mathbf{q}$ and frequency ω [14]

$$Q(\mathbf{q},\omega) = 2\pi\omega V^2\sum_{\mathbf{k}}(n(k)-n(k+q))\delta(\epsilon_{\mathbf{k}+\mathbf{q}}-\epsilon_{\mathbf{k}}-\hbar\omega) \quad , \tag{4.71}$$

which may be written as[15]

$$Q(\mathbf{q},\omega) = \omega V^2 \Im\chi_o(\mathbf{q},\omega) \tag{4.72}$$

This is just as we should have expected: the imaginary part of the response function gives the energy dissipated by the perturbation, just as for a classical oscillator.

Correlation function

There is another important quantity related to the response function, which is the correlation function in the ground state. A classical correlation function for the density relates the density at one point in space and time to a nearby point, i.e.

$$g(\mathbf{r},t;\mathbf{r}',t') =<\rho(\mathbf{r},t)\rho(\mathbf{r}'t')> - <\rho>^2 \quad . \tag{4.73}$$

The subtraction of the average density squared means that g measures only the correlation between particles: we expect it to decay at large distance and large time differences in a fluid. If the system is in equilibrium, it can depend only on the time difference $t-t'$, and

---

[13] Again, we add the complex conjugate of V so that we get both absorption and emission

[14] The two terms are combined by the change of variables $k \to k+q$ in one of them, as before

[15] Using, of course, the correct analytic continuation: $\lim_{\eta\to 0} 1/(x-i\eta) = 1/x + i\pi\delta(x)$.

in a homogeneous, translationally invariant system, only on the spatial difference $\mathbf{r}-\mathbf{r}'$. Thus we can just look at the fourier transform

$$S(\mathbf{q},\omega) =< \rho(\mathbf{q},\omega)\rho(-\mathbf{q},-\omega) > \tag{4.74}$$

which is sometimes called the structure factor.

In a quantum system we need to calculate averages of operators, say in some stationary state k. We shall also immediately move to fourier space (in frequency) and ask about

$$\begin{aligned} & \frac{1}{2} < k|\hat{\rho}(\omega)\hat{\rho}(\omega') + \hat{\rho}(\omega')\hat{\rho}(\omega)|k > \\ = & \frac{1}{2}\sum_l \left[< k|\hat{\rho}(\omega)|l >< l|\hat{\rho}(\omega')|k > + < k|\hat{\rho}(\omega')|l >< l|\hat{\rho}(\omega)|k >\right] \quad , \end{aligned} \tag{4.75}$$

where we have introduced a complete set of states, using the completeness relation

$$\sum_l |l >< l| = 1 \quad . \tag{4.76}$$

The time-dependence of an operator just means that its matrix elements must be calculated using time-dependent wavefunctions, viz.

$$\begin{aligned} < l|\hat{\rho}(\omega)|k > &= \int dt e^{i\omega t} e^{-i(\epsilon_k-\epsilon_l)t/\hbar} < \hat{\rho} >_{lk} \\ &= 2\pi < \hat{\rho} >_{lk} \delta(\omega-\omega_{kl}) \quad . \end{aligned} \tag{4.77}$$

Again, with a certain amount of algebra, one can evaluate all of this and show that

$$\begin{aligned} \frac{1}{2} < k|\hat{\rho}(\omega)\hat{\rho}(\omega') &+ \hat{\rho}(\omega')\hat{\rho}(\omega)|k > \\ &= \delta(\omega+\omega')\pi\sum_l | < \hat{\rho} >_{lk} |^2 \left[\delta(\omega+\omega_{kl}) + \delta(\omega-\omega_{kl})\right] \end{aligned} \tag{4.78}$$

Applying this further to our problem where the states are plane waves, eventually one reaches the result

$$S(\mathbf{q},\omega) = \pi\hbar\sum_{\mathbf{k}}(n(\mathbf{k}) - n(\mathbf{k}+\mathbf{q}))\delta(\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} + \hbar\omega) \tag{4.79}$$

$$= \hbar\Im\chi_o(\mathbf{q},\omega) \quad . \tag{4.80}$$

4.3.2 Response functions and the fluctuation - dissipation theorem

The algebra has been a little extended, so here are the three results.

First we calculated directly the density response function $\chi_o = \frac{\partial<\rho>}{\partial V}$, with the result

$$\chi_o(\mathbf{q},\omega) = 2\sum_{\mathbf{k}} \frac{n(\mathbf{k}) - n(\mathbf{k}+\mathbf{q})}{(\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}}) - \hbar\omega - i\eta} \tag{4.81}$$

This is a formula that one could almost have guessed. Remember we are applying a perturbation with momentum $\mathbf{q}$ and frequency ω. Thus it can produce scattering only between the plane wave eigenstates $\mathbf{k}$ and $\mathbf{k}+\mathbf{q}$; in first order perturbation theory there is a resonant energy denominator[16]

$$\frac{1}{(\epsilon_{\mathbf{k}+\mathbf{q}}-\epsilon_{\mathbf{k}})+\hbar\omega} \tag{4.82}$$

The occupation factors $(n(\mathbf{k})-n(\mathbf{k}+\mathbf{q}))$ appear because we can only make transitions from below the Fermi surface to above it, which limits the allowed transitions. The factor of 2 is for spin.

We then discovered that the energy dissipated by applying a perturbation could be written

$$\begin{aligned} Q(\mathbf{q},\omega) &= 2\pi\omega V^2\sum_{\mathbf{k}}(n(\mathbf{k})-n(\mathbf{k}+\mathbf{q}))\delta(\epsilon_{\mathbf{k}+\mathbf{q}}-\epsilon_{\mathbf{k}}-\hbar\omega) && (4.83)\\ &= \omega V^2\Im\chi_o(\mathbf{q},\omega) \quad . && (4.84)\end{aligned}$$

The relation makes use of the one important piece of complex algebra we need in the course

$$\lim_{\eta\to 0^+}\frac{1}{x+i\eta}=\wp\frac{1}{x}-i\pi\delta(x) \quad . \tag{4.85}$$

This is the same result that you should be familiar with for a classical oscillator. Again the result is intuitive; because we are dealing with absorption or radiation of energy, the δ-function represents energy conservation. The process involves removing an electron from below the Fermi surface, and putting it in a state above the Fermi surface - with both momentum and energy conserved. Often this is viewed as the creation of an *electron-hole pair.*

Furthermore, we found that the correlation function is directly related to the imaginary part of the response function

$$\begin{aligned} S(\mathbf{q},\omega) &= \int dt e^{i\omega t} < \hat{\rho}_{\mathbf{q}}(t)\hat{\rho}_{\mathbf{q}}(0) > \\ &= \hbar\Im\chi_o(\mathbf{q},\omega) \quad . && (4.86)\end{aligned}$$

This correlation function is often referred to as the *dynamic structure factor.* This last result is no accident, but an example of the *fluctuation – dissipation theorem.* Although we have derived it for the specific case of the density fluctuations in a fermi liquid, the relationship between S and $\Im\chi$

---

[16] The signs arise because the Schrodinger wavefunctions have a time-dependence $\exp(-i\epsilon_{\mathbf{k}}t/\hbar)$ with an opposite sign to our chosen convention for Fourier transforms; this is the source of unending suffering. However, $\Re\chi_o$ is an even function of frequency, $\Im\chi_o$ is odd. η is an infinitesimal positive number, put here to make sure that causality is respected.

is general. Remarkably, note that it relates a response function (i.e. properties of a perturbed system) to quantum fluctuations in the unperturbed system.[17]

The relations between energy loss Q, structure factor S and susceptibility χ are general for any response function, though here we have just exhibited them for noninteracting fermions. These relations provide the foundations to interpret the results of many experimental probes of condensed matter systems. Directly related to the density response function of electrons are the optical conductivity $\sigma(\omega)$ and the EELS spectrum, and inelastic light scattering (Raman or X-ray). In all cases these are probes that interact by coupling to the fluctuations in the electronic charge density. The d.c. conductivity is just the static limit of $\sigma(\omega)$.

The density response of the nuclear motion (i.e. phonons) can be seen in inelastic neutron scattering, and also in optical absorption or inelastic light scattering (near $q \approx 0$).

The electronic spin density fluctuations are observed by in inelastic neutron scattering, because the neutron has a magnetic moment. The spin-density correlation function may be constructed from a formula just like Eq. (4.86), but with the spin density operator replacing the number density operator; since the spin density couples to a magnetic field $\mathbf{H}$ through a term in the Hamiltonian $\mathbf{H} \cdot \mathbf{S}$, then the dynamic spin response function $\chi_{spin} = M/H$ will be related to the spin density correlation function. The static limit $\omega \to 0$ yields the Pauli paramagnetic susceptibility.

4.3.3 Screening and the dielectric function

Unfortunately, electrons interact, If we want to use the above results to calculate the density response of charged electrons, we must be careful about the effects of the long range Coulomb interaction. As we discovered earlier, screening in a metal arises because the *bare* or *external* potential induces a change in the charge density, and that charge density itself causes a change in the potential; the total potential that acts on the carriers in the metal is the sum of the two, and they largely cancel.

We can incorporate the screening effect by replacing the bare potential by the screened potential, and because we are dealing with negatively charge electrons, we have to be careful with signs. We are here just going to repeat the arguments we gave in Sec. 2.2.5, but now using a dynamical potential, rather than a static one.

The screened potential is given by

$$V_{scr}(\mathbf{q}, \omega) = V_{ext}(\mathbf{q}, \omega) - \rho(\mathbf{q}, \omega) v_{ee}(q) \tag{4.87}$$

---

[17] It also applies to classical thermal systems, and the result is modified at finite temperature: $S = \hbar \Im chi \coth(\hbar\omega/2T)$

Note that we have used here the (positive) density ρ, rather than the (negative for electrons) charge density $-|e|\rho$.

Following the earlier calculation, we now write

$$\rho = \chi_o V_{scr} \tag{4.88}$$

with the argument $(\mathbf{q}, \omega)$ suppressed for clarity, and we obtain

$$\rho = \frac{\chi_o V_{ext}}{1 + \chi_o v_{ee}} \tag{4.89}$$

An alternative way of stating this result is the ratio between the screened and the external potential

$$\frac{V_{scr}}{V_{ext}} = \frac{1}{\epsilon(\mathbf{q}, \omega)} = \frac{1}{1 + v_{ee}(\mathbf{q})\chi_o(\mathbf{q}, \omega)} \quad . \tag{4.90}$$

This gives us the formula for the dielectric function in what is called, for obscure historical reasons, the *Random Phase Approximation* or RPA.

$$\epsilon^{RPA}(\mathbf{q}, \omega) = 1 + \frac{8\pi e^2}{q^2} \sum_{\mathbf{k}} \frac{n_{\mathbf{k}} - n_{\mathbf{k+q}}}{\epsilon_{\mathbf{k+q}} - \epsilon_{\mathbf{k}} + \omega + i\eta} \quad . \tag{4.91}$$

It is much better understood as a *self-consistent field* theory, where the density response self-consistently adjusts to the external and its self-generated potential. It is the dynamical generalisation of Thomas-Fermi theory.

4.3.4 Properties of the RPA density response function

We are now back at the point where we can do physics again. Of course, we need to calculate the behaviour of χ_o, which is just algebra. Here are the main results, and the details are below, in small type. The dielectric function behaves very differently in the limits $\mathbf{q} \rightarrow 0$, (with $\omega \neq 0$), and $\omega \rightarrow 0$ (with $\mathbf{q} \neq 0$).

In the static limit $\omega = 0$, we have

$$\epsilon^{RPA}(\mathbf{q}, 0) = 1 + \frac{q_{TF}^2}{q^2} F(q/2k_F) \tag{4.92}$$

where the function $F(x) \rightarrow 1$ as $x \rightarrow 0$ (so that we recover the Thomas Fermi result), and the function itself is sketched in Fig. 4.10.

At $q = 2k_F$, χ_o is singular in one dimension, and has discontinuous derivatives in two and three dimensions. A singular static response in 1D means that a one-dimensional metal is always unstable - this is called the Peierls transition, and is the subject of question 4.5. The derivative singularities in higher dimensions mean that in real space $\chi_o(r)$ *oscillates* as a function of distance, along with the decay - these oscillations, called Friedel

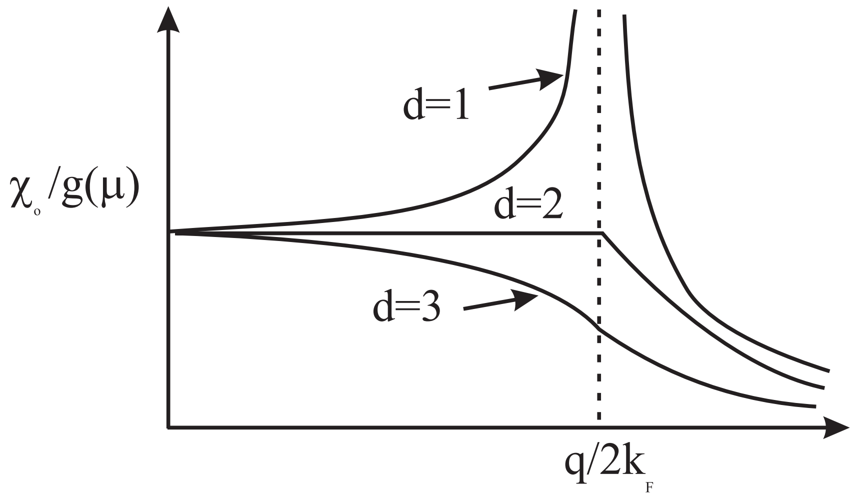


Figure 4.10: Sketch of the static susceptibility $\chi_o(\mathbf{q},0)$ as a function of $q/2k_F$ in one, two and three dimensions

oscillations, are shown in Fig. 4.11, and are important in mediating magnetic interactions between local moments in a metal. It can be shown that the induced charge density produced by an external point charge Q decays (in three dimensions) as

$$\rho_{ind}(r) \propto \frac{Q}{r^3}\cos(2k_F r) \quad , \tag{4.93}$$

In the long-wavelength limit, but at nonzero frequency, we find

$$\epsilon^{RPA}(0,\omega) = 1 - \frac{\omega_p^2}{\omega^2} \quad , \tag{4.94}$$

which is the same as the classical result we got in our discussion of plasmons.

In both the static, and in the long-wavelength limits, the dielectric function is real; an imaginary part can exist only when the potential can induce real transitions between states, which is confined to a bounded region in $(\mathbf{q},\omega)$-space shown in Fig. 4.12.

A more detailed discussion (and calculation) of these results is given below.

Static response

Let us first look at the behaviour at $\omega = 0$, where the response function is purely real.

$$\chi_o(\mathbf{q}) = 2\sum_{\mathbf{k}} \frac{n_{\mathbf{k}} - n_{\mathbf{k+q}}}{\epsilon_{\mathbf{k+q}} - \epsilon_{\mathbf{k}}} \quad . \tag{4.95}$$

For $\mathbf{q} \to 0$, we have

$$\chi_o \quad = \quad -2\sum_{\mathbf{k}} \frac{\mathbf{q}\cdot\frac{\partial n_{\mathbf{k}}}{\partial \mathbf{k}}}{\mathbf{q}\cdot\frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}}}$$

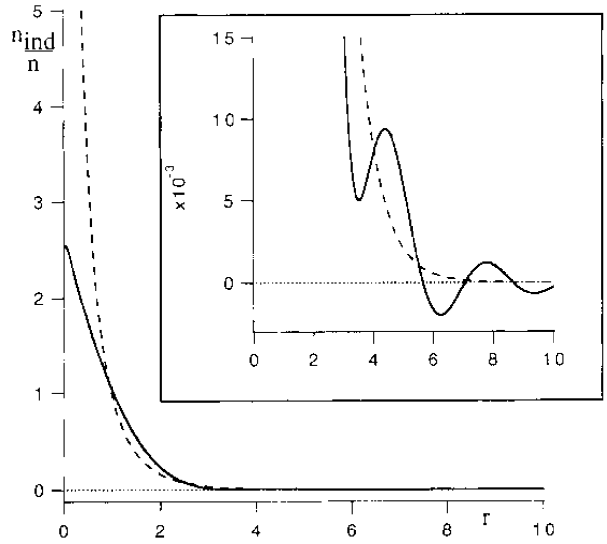


Figure 4.11: Behaviour of the induced charge density when a point charge $+|e|$ is inserted at the origin. The dashed line is the Thomas-Fermi approximation, which has the unphysical singularity at the origin; the solid line is the RPA response, finite at the origin, and showing Friedel oscillations at large distances.

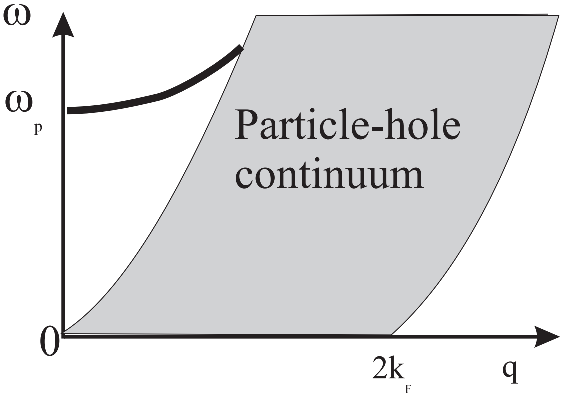


Figure 4.12: . Regions of (q, ω) space where the dielectric function and density response function have different characters. The imaginary part is non-zero only in the hatched region, defined by Eq. (4.107). Along the frequency axis ($q \approx 0$), we have $\chi_o = -\frac{nq^2}{m\omega^2}$ and $\epsilon = 1 - \omega_p^2/\omega^2$; along the momentum axis, we have ($q \ll 2k_F$) $\chi_o = g(\mu)$ and $\epsilon = 1 + q_{TF}^2/q^2$.

$$
\begin{aligned}
&= -2\sum_{\mathbf{k}} \frac{(\mathbf{q}\cdot\frac{\partial\epsilon_{\mathbf{k}}}{\partial\mathbf{k}})\frac{\partial n_{\mathbf{k}}}{\partial\epsilon_{\mathbf{k}}}}{\mathbf{q}\cdot\frac{\partial\epsilon_{\mathbf{k}}}{\partial\mathbf{k}}} \\
&= -\int d\epsilon g(\epsilon)\frac{\partial n(\epsilon)}{\partial\epsilon} \qquad (4.96)
\end{aligned}
$$

where we have made use of the fact that the occupancy depends only on the energy to transform the sum over momentum into an integral over energy; $g(\epsilon)$ is the density of states in energy that we derived in chapter 2. Now $n(\epsilon)$ is of course a step function at $\epsilon = \mu$, so we get

$$\chi_o(\mathbf{q}) = g(\mu) \qquad (4.97)$$

This result gives for the RPA dielectric function in the long-wavelength limit the same result we derived earlier in the Thomas-Fermi theory:

$$\epsilon(\mathbf{q}\to 0, \omega = 0) = 1 + \frac{4\pi e^2 g(\mu)}{q^2} = 1 + \frac{q_{TF}^2}{q^2} \quad . \qquad (4.98)$$

Larger momentum is more tricky, and we just quote the result (in three dimensions):

$$\epsilon^{RPA}(\mathbf{q}, 0) = 1 + \frac{q_{TF}^2}{q^2}F(q/2k_F) \quad , \qquad (4.99)$$

where $F(x)$ is a function we have seen before

$$F(x) = \frac{1}{2} + \frac{1-x^2}{4x}\ln\left|\frac{1+x}{1-x}\right| \quad . \qquad (4.100)$$

which decays monotonously with increasing argument.

We noted earlier that the function $F(x)$ has a logarithmic singularity in the derivative at $q = 2k_F$. This arises because if $q < 2k_F$, it is possible to find contributions to the integral where $\epsilon_{\mathbf{k}+\mathbf{q}} = \epsilon_{\mathbf{k}}$, but for $q > 2k_F$ this is not allowed. This subtle feature has important consequences, in the form of Friedel oscillations. If one goes from momentum space into real space, any sharp features in q will lead to *oscillations* in r. The slope singularity means that the induced charge density due to the screening of a local potential will have oscillations in space with wavevector $2k_F$; for a Coulomb potential of charge Q for example one gets

$$\rho_{ind}(\mathbf{r}\to\infty) \approx \frac{Q}{r^3}\cos(2k_F r) \quad . \qquad (4.101)$$

Such oscillations also occur in the spin density when a local exchange (i.e. spin-dependent) potential is applied to an electron gas. These spin density oscillations then generate the Ruderman-Kittel-Kasuya-Yoshida (RKKY) interactions between localised moments at a distance.

The singularity at $q = 2k_F$ becomes more pronounced in lower dimensions. In Fig. 4.10 we sketch the behaviour of the static susceptibility in one, two, and three dimensions. In 2D, there is a cusp, and in 1D a singularity. The one-dimensional singularity implies that the metallic state is unstable, of course.

Dynamic response

By doing the relabelling $\mathbf{k}+\mathbf{q}\to -\mathbf{k}$ in the $n(\mathbf{k}+\mathbf{q})$ term, the dynamic density response function can be rewritten[18]

$$\chi_o(\mathbf{q},\omega) = \frac{4}{V}\sum_{\mathbf{k}} n_{\mathbf{k}}\frac{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}}}{(\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}})^2 - (\hbar\omega + i\eta)^2} \qquad (4.102)$$

---

[18] We put back in the factor of 1/V before the momentum sums, and reintroduce $\hbar$

Clearly, if $\mathbf{q}$ is small and ω finite, we can perform an expansion of the denominator in powers of $(\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}})/\omega$; we shall just keep the first term, which is

$$\begin{aligned}
\chi_o(\mathbf{q},\omega) &\approx -\frac{4}{V}\sum_{\mathbf{k}} n_{\mathbf{k}} \frac{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}}}{(\hbar\omega + i\eta)^2} \\
&= -\frac{4}{V}\sum_{\mathbf{k}} n_{\mathbf{k}} \frac{\frac{\hbar^2}{2m}(2\mathbf{k}\cdot\mathbf{q} + q^2)}{(\hbar\omega + i\eta)^2} \\
&= -\frac{4}{V}\frac{\hbar^2 q^2}{2m(\hbar\omega + i\eta)^2}\sum_{\mathbf{k}} n_{\mathbf{k}} \\
&= -4\frac{\hbar^2 q^2}{2m(\hbar\omega + i\eta)^2}\frac{n}{2} \\
&= -\frac{nq^2}{m\omega^2} \quad . \qquad (4.103)
\end{aligned}$$

Note that the angular average kills the term linear in q, and in the next to last line we used $(1/V)\sum_{k<k_F} = n/2$.

We have in fact seen this result before, but in a different guise: substitute Eq. (4.103) into the formula for the dielectric function

$$\epsilon^{RPA}(\mathbf{q} \to 0, \omega) = 1 - \frac{4\pi n e^2}{m\omega^2} \qquad (4.104)$$

which we derived earlier in the context of plasmons and the energy loss function. The next order terms in Eq. (4.103) are of order $(v_F q/\omega)^2$ and give an upward dispersion to the plasmon, shown in Fig. 4.12.

This expansion is valid for any frequency ω provided $(v_F q/\omega) \ll 1$, and yields a dielectric function which diverges to $-\infty$ as $\omega \to 0$; this should be contrasted with the static dielectric function which diverges to $+\infty$ as $q \to 0$. In order to understand just how these two opposing singularities resolve themselves we need to look at the *imaginary* part of the response function, hitherto ignored,

$$\Im\chi_o(\mathbf{q}\omega) = 2\pi\sum_{\mathbf{k}}(n_{\mathbf{k}} - n_{\mathbf{k}+\mathbf{q}})\delta(\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} - \omega) \quad . \qquad (4.105)$$

Remember that the imaginary part of the response function refers to the energy dissipated by real transitions induced by the external potential; the δ-function is there to conserve energy, and the factor $(n_{\mathbf{k}} - n_{\mathbf{k}+\mathbf{q}})$ makes sure that the transition occurs from an occupied state to an unoccupied state. It looks at first sight as if it will give a series of sharp spikes, but in fact there is a continuum of values of $\mathbf{k}$ for which the condition can be satisfied. The energy difference between initial and final states is

$$\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}(q^2 + 2\mathbf{k}\cdot\mathbf{q}) \qquad (4.106)$$

and thus, depending on the angle between $\mathbf{k}$ and $\mathbf{q}$

$$\frac{\hbar^2}{2m}(q^2 - 2qk_F) < \epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}} < \frac{\hbar^2}{2m}(q^2 + 2qk_F) \quad , \qquad (4.107)$$

and therefore only for frequencies inside this range are excitations possible. These two boundaries are shown in Fig. 4.12.

The excitations that can be made in the system consist either of collective oscillations (the plasmon) or of continuum “single-particle”-like oscillations. The presence of the latter explains the dramatic sign changes in ϵ that we referred to above. Imagine that we could resolve the discrete transitions corresponding to individual particle-hole transitions (this

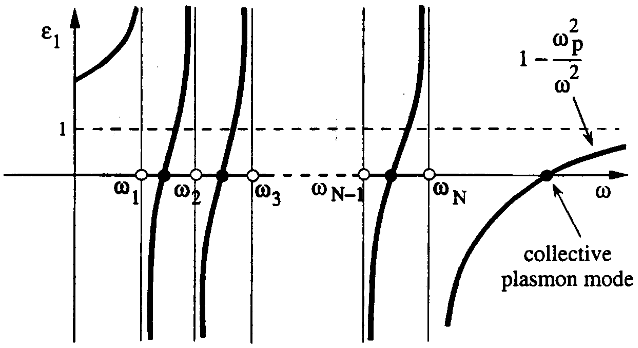


Figure 4.13: Behaviour of the real part of the dielectric function sweeping through a sequence of closely space poles, followed by an isolated pole at high frequencies – the plasmon

is correct in principle, because the momenta are all quantised inversely to the length of the system). Then at finite $\mathbf{q}$, and varying ω we can plot the real part of ϵ as it sweeps through the sequence of closely spaced poles, as in Fig. 4.13. ϵ must cross zero many times in the continuum, but there is a single isolated zero above the continuum, which is the collective mode.

Questions on chapter 4

Qu.4.1 LCR circuit

This question on electrical circuits is to remind you of the basic ideas of a dynamical response function, resonance, and dissipation.

An electrical circuit consists of an inductance L, resistance R and capacitance C in series, driven by a voltage source $V(t) = V_o \cos(\omega t)$. Show that the charge $Q(t)$ on the capacitor satisfies the equation

$$L\ddot{q} + R\dot{q} + q/C = V(t) \quad , \tag{4.108}$$

and use it to define the complex susceptibility from

$$q(\omega) = \chi(\omega)V(\omega) \tag{4.109}$$

Show that the forced solution of this equation is

$$q(t) = \frac{V_o \cos(\omega t - \phi)}{\left[\left(-\omega^2 L + \frac{1}{C}\right)^2 + (\omega R)^2\right]^{\frac{1}{2}}} \quad , \tag{4.110}$$

where

$$\tan(\phi) = \frac{\omega R}{\omega^2 L - 1/C} \quad . \tag{4.111}$$

Show that the mean rate of power dissipation is

$$W = \frac{1}{2} \frac{\omega V_o^2 \sin(\phi)}{\left[\left(-\omega^2 L + \frac{1}{C}\right)^2 + (\omega R)^2\right]^{\frac{1}{2}}} \quad . \tag{4.112}$$

Sketch the real and imaginary parts of χ as a function of frequency, for the cases $Q \ll 1$, $Q \approx 1$ and $Q \gg 1$, where $Q = (1/R)(L/C)^{\frac{1}{2}}$ is the "quality factor".

Where are the poles of χ in the complex ω plane?.

It is of course more usual to work with the complex impedance $Z = 1/(-i\omega\chi)$, *but this is a little obscuring because then the equation of motion for the current has a source term* $\dot{V}$.

Qu.4.2 Landau theory of phase transitions

An expansion of the free energy as a power series in a collective order parameter is called a Landau or Ginzburg-Landau expansion. It is very commonly used as a simple mean-field description of a phase transition. Here we use it to describe a ferro-electric.

A ferroelectric crystal is one that supports a macroscopic polarisation P – that usually arises because the underlying crystal structure does no have inversion symmetry. However, as temperature or pressure is changed, the crystal may recover the inversion symmetry. This can be modelled by

Landau's theory of second order phase transitions, where we postulate a form for the Free energy density (per unit volume)

$$\mathcal{F} = \frac{1}{2}aP^2 + \frac{1}{4}bP^4 + \frac{1}{6}cP^6 + ... \tag{4.113}$$

where the coefficient $a = a_o(T - T_c)$ is temperature dependent, and all the other coefficients are constant. Although the polarisation P is of course a vector, we assume that it can point only in a symmetry direction of the crystal, and so it is replace by a scalar.

(a) Assume that $b > 0$ and $c = 0$. Use Eq. (4.113) to determine the form for the equilibrium $P(T)$.

(b) Including in $\mathcal{F}$ the energy of the polarisation coupled to an external electric field E, determine the dielectric susceptibility $\chi = \frac{\partial P}{\partial E}$ both above and below the critical temperature.

(c) Sketch curves for $P(T)$, $\chi^{-1}(T)$, and $\chi(T)$.

(d) In a different material, the free energy is described by a similar form to Eq. (4.113), but with $b < 0$ and $c > 0$. By sketching $\mathcal{F}$ at different temperatures, discuss the behaviour of the equilibrium polarisation and the linear susceptibility, contrasting the results with those found in (c).

Qu.4.3 Reflectivity of metals

The phase velocity of light in a conducting medium is the speed of light divided by the complex dielectric constant $N(\omega) = \epsilon(\omega)^{1/2}$ where we may use for ϵ the Drude result

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega/\tau} \quad . \tag{4.114}$$

In a good Drude metal, we have $1/\tau \ll \omega_p$.

Sketch curves of

(a) $\Re\sigma(\omega)$,

(b) $\Re\epsilon(\omega)$,

(c) $\Im(1/\epsilon(\omega)$.

Consider a light wave with the electric field polarised in the $x-$direction at normal incidence from the vacuum on a good Drude metal (with $1/\tau \ll \omega_p$) occupying the region $z > 0$. In the vacuum ($z < 0$), the incident E_1 and reflected E_2 waves give rise to a field

$$E_x = E_1 \exp(i\omega[z/c - t]) + E_2 \exp(-i\omega[z/c + t]) \quad , \tag{4.115}$$

whereas in the medium, the electric field is

$$E_x = E_0 \exp(i\omega[N(\omega)z/c - t]) \quad . \tag{4.116}$$

Matching the electric and magnetic fields on the boundary, show that

$$\begin{aligned} E_0 &= E_1 + E_2 \quad , && (4.117)\\ NE_0 &= E_1 - E_2 \quad , && (4.118) \end{aligned}$$

and hence show that the reflection coefficient satisfies

$$R = \left|\frac{E_2}{E_1}\right|^2 = \left|\frac{1-N}{1+N}\right|^2 \quad . \tag{4.119}$$

Using the Drude formula above, show that

$$\begin{aligned} R &\approx 1 - 2\left(\frac{\omega}{2\pi\sigma(0)}\right)^{1/2} && \text{for } \omega \ll 1/\tau && (4.120)\\ &\approx 1 - \frac{2}{\omega_p \tau} && \text{for } 1/\tau \ll \omega \ll \omega_p && (4.121)\\ &\approx 0 && \text{for } \omega_p \ll \omega && (4.122) \end{aligned}$$

and sketch the reflectivity $R(\omega)$.

Deducing the complex conductivity by measuring the reflectivity is a standard experimental technique. To get both real and imaginary parts from a measurement of only $|R|$ (rather than the complex R) requires employment of the Kramers-Krönig relations.

Qu.4.4 Phonons

From Eq. (4.8) construct $\Im\chi$ in the limit that $\gamma \to 0$. Use the Kramers Krönig relation to then reconstruct $\Re\chi$ from $\Im\chi$ in the same limit.

Qu.4.5 Screened Coulomb interaction

Consider a nucleus of charge Z producing an potential

$$V_{ext}(\mathbf{q}) = -\frac{4\pi Ze^2}{q^2}$$

Using the long-wavelength limit of the dielectric function, show that the screened potential satisfies

$$V_{scr}(\mathbf{q}=0) = -\frac{2}{3}\Omega E_F$$

where Ω is the volume of the unit cell, and E_F is the Fermi energy for Z free electrons per unit cell.

Qu.4.6 Peierls transition

By rewriting the term containing $n_{\mathbf{k}+\mathbf{q}}$ (replace $\mathbf{k}+\mathbf{q} \to -\mathbf{k}'$ and then relabel $\mathbf{k}'$ as $\mathbf{k}$), show that the static density response function can be written

$$\chi_o(\mathbf{q}, 0) = 2\sum_{k<k_F} \frac{1}{\epsilon_{\mathbf{k}+\mathbf{q}} - \epsilon_{\mathbf{k}}} \quad . \tag{4.123}$$

In one dimension, make a linear approximation to the electronic dispersion near k_F, i.e. $\epsilon_{\mathbf{k}} = v_F|k|$, and consider the response for $q = 2k_F + p$, where $p \ll 2k_F$. By considering terms in the sum over k near $k \approx -k_F$, show that

$$\chi_o(2k_F + p) \approx \frac{1}{2\pi v_F} \ln \left| \frac{2k_F}{p} \right| \quad . \tag{4.124}$$

Explain why this result suggests that a one-dimensional metal will be unstable to a lattice distortion with wavevector $2k_F$.

This result is exactly analogous to the nonanalytic behaviour we saw in Question 2.5. There we found that the energy gain from opening a gap U at the chemical potential was $\approx U^2 \ln|E_F/U|$, *when we chose the periodicity of the potential to be exactly* $2k_F$. *Remember that linear response theory will predict that*

$$\mathcal{U} = -\rho V_{ext} + \frac{1}{2}\chi^{-1}\rho^2 \tag{4.125}$$

which after minimisation, yields

$$\mathcal{U}_{min} = -\frac{1}{2}\chi V_{ext}^2 \tag{4.126}$$

In our linear response calculation, $\chi_o(2k_F)$ *is singular, indicating as usual the failure of non-degenerate perturbation theory, whereas the exact calculation done earlier showed that* $\chi_0(2k_F) \propto \ln|E_F/V_{ext}|$ *– a result that could perhaps have been guessed from Eq. (4.124).*

Qu.4.7 Optical properties

Discuss why, at optical frequencies, glass is transparent, and silver is shiny, while graphite appears black, and powdered sugar is white.

Qu.4.8 Metals and insulators

Explain the differences between a metal and an insulator. Your discussion should include reference to: single particle properties; screening of the Coulomb interaction; optical properties; and electrical and thermal properties.

Chapter 5

Magnetism

We have already gained quite some familiarity with the origins of magnetic interactions either from Hartree-Fock theory in the electron gas, or from the point of view of localised exchange interactions on transition metal ions. In this chapter we will pull these points of view together, and also look at magnetism as a collective phenomenon, using the tools of the last chapter.

5.1 The origin of local magnetic moments

Strongly-bound core states in atoms acquire magnetic moments because of interactions between ions as we discussed in Sec. 3.3. The general rule is that many-particle wavefunctions that are built out of *orthogonal* orbitals have a tendency to spin-alignment. This is called Hund's rule, and can be seen simply within the Hartree Fock approximation.

As a simple example, consider a model atom with two orbitals of single-particle energies $E_{A,B}$ in which we wish to accommodate two electrons (Fig. 5.1). Often they will be of similar, or identical energies, but for definiteness, let's take $E_A < E_B$. Within Hartree-Fock (see Eq. (2.95)) in this simple

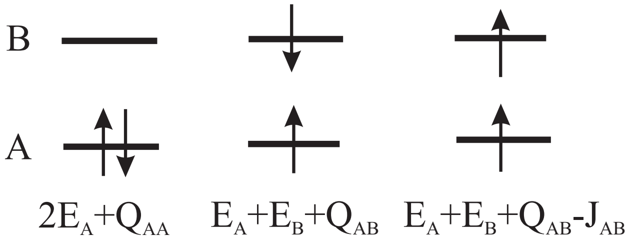


Figure 5.1: Possible low energy configurations of two electrons in a two-level atom

model, there are the following interaction terms: the direct (Hartree) terms

$$\begin{aligned} Q_{AA} &= < AA|\frac{e^2}{r_{12}}|AA > \\ Q_{AB} &= < AB|\frac{e^2}{r_{12}}|AB > \\ Q_{BB} &= < BB|\frac{e^2}{r_{12}}|BB > \end{aligned} \tag{5.1}$$

and the exchange energy

$$J_{AB} =< AB|\frac{e^2}{r_{12}}|BA > \tag{5.2}$$

that operates only betgween configurations where the single particle states are occupied with identical spin. Since the orbitals are assumed orthogonal, then $Q_{AA} \approx Q_{BB} \gg Q_{AB} \approx J_{AB}$

The three configurations have the energy shown in Fig. 5.1; two are singlets, and the last is a triplet. The triplet configuration will have the lowest energy if

$$Q_{AA} - Q_{AB} + J_{AB} > E_B - E_A \tag{5.3}$$

which is very commonly the case when Q_{AA} (which we have elsewhere modelled as the Hubbard U parameter) is large.

This simple example reflects a general phenomenon. Remember that we can factorise the wave function into a product of orbital and spin components. Since the total wavefunction must be antisymmetric, if we choose the spin wavefunction to be symmetric under exchange of coordinates (which enforces a triplet: one of $|\uparrow\uparrow>; |\downarrow\downarrow>; 2^{-1/2}(|\uparrow\downarrow> + |\downarrow\uparrow>)$)then the spatial part of the wavefunction is *anti-* symmetric. The singlet state $2^{-1/2}(|\uparrow\downarrow> - |\downarrow\uparrow>)$ is antisymmetric in spin space, and therefore the real space wavefunction must be symmetric. An antisymmetric wavefunction must have nodes whenever two spatial coordinates are equal: $\psi(....,r_i = r, ...r_j = r, ...) = 0$. So it is then clear that the particles stay farther apart in an antisymmetrised state than in a symmetric state, and because of the Pauli principle an antisymmetric wavefunction (which will generally have high spin) has lower energy.

The physical reason for the existence of local moments on atoms is then a combination of the Pauli principle together with repulsive interactions between ions. If we consider, say, d-levels in an ion, since the d-states are degenerate, we shall always get high spin configurations. However, in the environment of a solid, the d-levels are split because the atom is no longer in a potential of spherical symmetry. If this *crystal field splitting* is large enough, then the condition of Eq. (5.3) will not be satisfied, and the orbitals will be filled one after another - and generally the atom will have low spin.

5.1.1 Spin Hamiltonian and the Heisenberg Model

If we focus only on those states with the same orbital occupancy, they differ only by their spin degrees of freedom - for example the two rightmost states in Fig. 5.1. If $Q_{AA} \approx Q_{BB} \approx U$ is large, only the fluctuations between these two spin configurations will be important. The spatial degrees of freedom of the wavefunction are frozen in place, and we can work in models with spin degrees of freedom alone. Clearly, this is only going to be justified in an insulator [1] but such a situation is quite common.

We then need to write down a Hamiltonian for the spins $\mathbf{S}_A$, $\mathbf{S}_B$. First note that the total spin $\mathbf{S}$ is given by

$$\mathbf{S}^2 = (\mathbf{S}_A + \mathbf{S}_B)^2 = (\mathbf{S}_A)^2 + (\mathbf{S}_B)^2 + 2\mathbf{S}_A \cdot \mathbf{S}_B = \frac{3}{2} + 2\mathbf{S}_A \cdot \mathbf{S}_B \tag{5.4}$$

where we have used for a spin-half state $\mathbf{S}_i^2 = \frac{1}{2}(\frac{1}{2}+1) = \frac{3}{4}$. Now $\mathbf{S}^2$ has the eigenvalue $S(S+1)$ in a state of spin S, it then follows that the eigenvalue of $\mathbf{S}_A \cdot \mathbf{S}_B$ is $-\frac{3}{4}$ in the singlet state ($S = 0$), and $\frac{1}{4}$ in the triplet state. This allows us to cast the spin Hamiltonian as

$$H_{spin} = \frac{1}{4}(E_s + 3E_t) - (E_s - E_t)\mathbf{S}_A \cdot \mathbf{S}_B \tag{5.5}$$

with $E_{s,t}$ the energies of the singlet and triplet states. If we drop the constant term by redefining the zero of energy, we have then

$$H_{spin} = -J_{AB}\mathbf{S}_A \cdot \mathbf{S}_B \tag{5.6}$$

where we see that the model will favour parallel spins if $J > 0$, and antiparallel spins if $J < 0$.

5.2 Types of magnetic interactions

The existence of magnetic moments locally on atoms is only part of the story of magnetism, because to obtain an ordered magnetic ground state, the moments between neighbouring atoms must be coupled. There are a number of different ways that these couplings can arise. The net effect of the couplings between neighbours can in the end usually be parametrised quite simply, just in terms of pairwise interactions between spins *on neighbouring sites*, viz.

$$H_{spin} = -\sum_{ij} J_{ij}\mathbf{S}_i \cdot \mathbf{S}_j \tag{5.7}$$

This is the Heisenberg model, which now couples the total spin on an atom (which will be determined by the solution of the atomic problem Eq. (5.6))

---

[1] Or where the conduction electrons are only weakly coupled to spin degrees of freedom in the core

to that of its neighbours. Notice that the coupling only depends on the relative orientation of the spins, and not on their absolute direction relative to the crystal lattice. When the angular momentum of the ion contains an orbital part as well as a spin part, then the spin Hamiltonian will include a piece that depends on the *absolute* spin direction.

5.2.1 Dipolar Interaction

The first idea might just be that the moments could couple via through the magnetic fields they generate. However, this is very small: the energy of interaction of two magnetic dipoles of strength m at a distance r is of order $\mu_o m^2/4\pi r^3$. Putting in a magnetic moment of order a Bohr magneton, we get

$$U_{dipolar} \approx \frac{\mu_o}{4\pi}(\frac{e\hbar}{2m})^2\frac{1}{r^3} \approx \pi\alpha^2(\frac{a_{Bohr}}{r})^3 Ryd. \tag{5.8}$$

where $\alpha \approx 1/137$ is the fine structure constant. At typical atomic separations of $2nm$, this is about 4×10^{-5} eV, or less than a degree Kelvin. As always, magnetic interactions are over-ruled by charge interactions, and such energy scales are rarely important. [2]

Dipolar terms do play an important role in the domain structure of magnets, once the spins have already been aligned by other means.

5.2.2 Direct, itinerant, and super-exchange

The intra-atomic exchange interaction we discussed in Sec. 5.1 is an example of *direct exchange*, because it comes from interactions between *overlapping* orbitals. When the orbitals concerned are orthogonal, J is *positive* in sign, i.e. the lowest energy state is a triplet. However, if the overlapping orbitals are not orthogonal – as will happen between two orbitals between neighbouring atoms – the interaction may be of a *negative* sign, so the lowest energy is a singlet.

In molecular H_2, for example, the singlet is *lower* than the triplet – because although the *symmetric* spatial wavefunction causes the two electrons to be closer together than in the antisymmetric state (so increasing the repulsive inter-electronic Coulomb energy), this is counterbalanced by the spatial wavefunctions having charge density maxima located in the bond between the ions. So covalent bonds are singlet pairs, and covalent solids are magnetically inert.

In d- and f- electron magnets, the direct overlap between the atomic magnetic states on different ions is usually very small. Magnetic interactions between such atoms arise when the magnetic interaction is mediated by an otherwise nonmagnetic electrons. There are two common cases. One is

---

[2] Interactions between electrical dipoles – as occurs in a ferroelectric – are not negligible. Here we get the same expression but without the factor of α^2

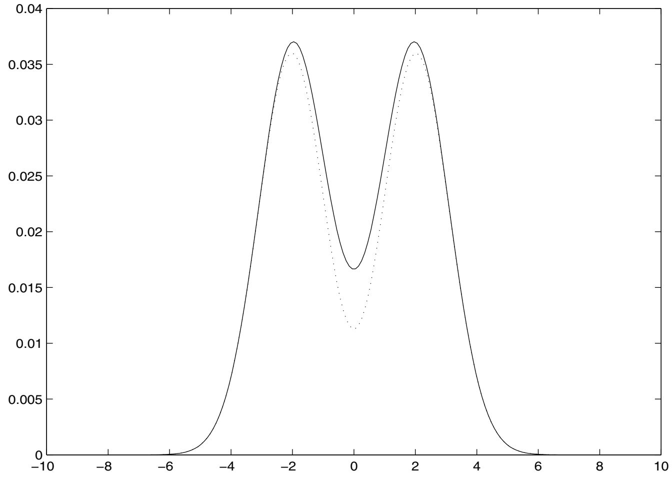


Figure 5.2: Schematic picture of the origin of the superexchange interaction. The figure shows the charge density for the wavefunctions in a singlet state (solid line) and a triplet state(dotted line)

where the two magnetic atoms are separated by a nonmagnetic ion with a closed shell, commonly in an insulating magnetic oxide where the oxygen ion is O^{2-}. This is called *superexchange*. The second case is in a magnetic d- or f-band metal, where there are itinerant s- or p-like electronic states. A magnetic interaction mediated by conduction electrons is called *itinerant exchange*.

Superexchange can be visualised in Fig. 5.2. We have in mind here a Mott insulating state, so that the direct overlap between the d-electrons on the nearest magnetic ions is too small to overcome the local Coulomb repulsion. Furthermore, because the mediating O ion is almost completely full, there is only a little overlap between the d-wavefunction and the oxygen neighbour. This small overlap is, however, just enough to generate an exchange interaction.

As usual, if we consider the spatial wavefunctions for the singlet and triplet states, they will be of the form

$$\psi = \frac{1}{\sqrt{2}}[\psi_A(1)\psi_B(2) \pm \psi_A(2)\psi_A(1)] \tag{5.9}$$

where the + sign goes with the spin singlet, and the − sign with the triplet. Because the triplet state has a node, it has less possibility for both electrons to spill over onto the O ion than does the singlet. Consequently the singlet state has a somewhat lowered kinetic energy (the electrons a little less localised) and this is enough to lower its energy below the singlet state. So for superexchange J is *negative* favouring antiparallel, or *antiferromagnetic* alignment of spins. Values of J for transition metal oxides range from a few tens of Kelvin to a few hundreds.

Itinerant exchange comes about in metals because the atomic moments produce a spin-polarisation of the itinerant cloud. If we have an ion of spin **S** embedded in the conduction electrons, one would expect that the local direct exchange will give rise to a contact interaction of the form

$$H_{int} = -J\mathbf{S}\cdot\mathbf{s}\delta(\mathbf{r}) \quad , \tag{5.10}$$

with **s** the conduction electron spin density, and J a direct exchange interaction. The spin density is not otherwise polarised, but the perturbation will induce a weak spin density modulation in the conduction cloud, which will of course decay away to zero at large distance from the ion. The induced spin density is just

$$s(\mathbf{r}) = J\chi_\sigma(\mathbf{r})S \tag{5.11}$$

using the definition of the spin susceptibility χ_σ. At a nearby lattice site (say **R**), the induced spin density caused by the polarisation of one atom interacts with the spin of another, and the energy is then

$$-J\mathbf{S}(\mathbf{R})\cdot\mathbf{s}(\mathbf{R}) = J^2\chi_\sigma(\mathbf{R})\mathbf{S}(\mathbf{R})\cdot\mathbf{S}(0) \quad , \tag{5.12}$$

Summing over all pairs of sites in the crystal we obtain

$$H_{RKKY} = -\sum_{ij} J^2\chi_\sigma(\mathbf{R}_{ij})\mathbf{S}(\mathbf{R}_i)\cdot\mathbf{S}(\mathbf{R}_j) \quad , \tag{5.13}$$

which is the RKKY (Ruderman-Kittel-Kasuya-Yoshida) interaction.

If the itinerant electrons can be approximately described as a free electron gas, then $\chi_\sigma = \chi_o$ derived for a density perturbation in the electron gas in chapter 4. (The interaction term can be thought of as a potential which has interactions of the opposite sign between up and down electrons.) We saw that χ_o is of order $N(\mu)$, the density of states per unit energy, and is positive for small r (ferromagnetism), but has Friedel oscillations at large distances of the form $\cos(2k_F r)/r^3$, so the interaction will then oscillate in sign. If the neighnouring ions are close, the Heisenberg interatomic exchange parameter is of order $J^2N(\mu)$ which can be large - up to 1 eV or so. The RKKY interaction is visualised in Fig. 5.3.

5.3 Stoner-Hubbard theory of itinerant magnetism

In the pictures of magnetism described in the last section, we relied on the atomic physics to generate the moment, and then were concerned only with the (weaker) interactions between the moments. However, it would be pedagogogically useful to be able to start with a picture of itinerant electrons in Bloch states, with the moments determined self-consistently by taking proper account of exchange and correlation effects. We have already addressed this partially when we looked at ferromagnetism in the electron gas within the Hartree-Fock approximation.

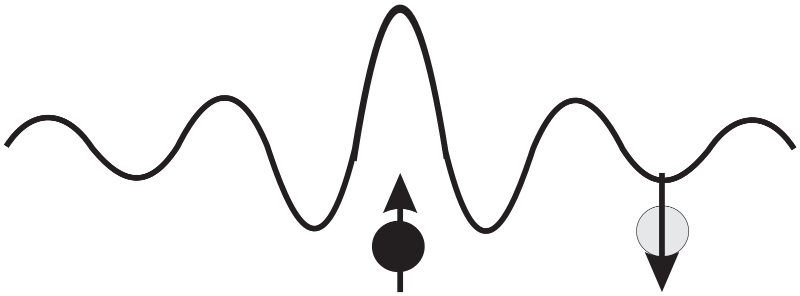


Figure 5.3: . In metals, a local moment will polarise the conduction electron spins, producing a spin density that decays away and oscillates in sign with period $1/2k_F$. The interaction of the induced spin density with a neighbouring local moment produces the RKKY interaction

5.3.1 Spin paramagnetism in metals

We will begin this section by reviewing the standard theory of Pauli paramagnetism. We consider a fermi gas with energy dispersion $\epsilon_{\mathbf{k}}$ in a magnetic field H. Because of the energy of the spin in a magnetic field, the spin-up and spin-down bands will be split (see Fig. 5.4), and have energies

$$\begin{aligned} \epsilon_{\mathbf{k}\uparrow} &= \epsilon_{\mathbf{k}} + \mu_B H \quad , \\ \epsilon_{\mathbf{k}\downarrow} &= \epsilon_{\mathbf{k}} - \mu_B H \quad . \end{aligned} \tag{5.14}$$

Since the chemical potential must be the same for both spins, there must be a transfer of carriers from the minority spin band to the majority

$$n_\uparrow - n_\downarrow = \mu_B H g(\mu) \tag{5.15}$$

where $g(\mu)$ is the density of states at the Fermi level[3]. The magnetisation is $M = \mu_B(n_\uparrow - n_\downarrow)$ which then gives us the static spin susceptibility

$$\frac{M}{H} = \chi_\sigma(\mathbf{q}=0, \omega=0) = \mu_B^2 g(\mu) \quad . \tag{5.16}$$

5.3.2 Ferromagnetism in the Stoner-Hubbard model

Now let us include in a very simple fashion the effect of interactions. We will turn to the Hubbard model again, where now in addition to the kinetic energy of the electrons, we put in an interaction

$$\hat{H}_{int} = U \sum_i \hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \quad , \tag{5.17}$$

---

[3] Obviously, we must assume that the splitting is small enough that the density of states can be taken to be a constant. We define $g(\mu)$ to be the density of states for both spins.

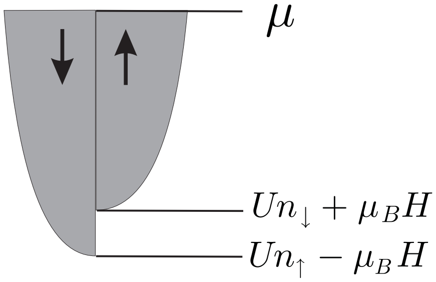


Figure 5.4: Spin-split bands in the Stoner approximation

where $\hat{n}_{i\sigma}$ is the number operator for the electron on lattice site i. In contrast to our Hartree-Fock treatment when we used the full long-range Coulomb interaction, we expect the effective interactions between electrons to be short-range due to screening, and they are parametrized here by the Hubbard interaction U. It is of course just the Coulomb energy of having two electrons on the same site — and notice that because of the Pauli principle they must already have opposite spins.

In a ferromagnet, the spin and charge are the same on every site, so $< \hat{n}_{i\downarrow} >= n_\downarrow = N_\downarrow/N$ and $< \hat{n}_{i\uparrow} >= n_\uparrow = N_\uparrow/N$, with $N = N_\uparrow + N_\downarrow$ the total number of electrons. In a simple mean field approximation, the interaction term is then just replaced by a potential $Un_\uparrow$ for the $\downarrow$ spin, and $Un_\downarrow$ for the $\uparrow$ spin. As in the last section, we also include an applied field, which then gives a Zeeman energy $-2\mu_B \mathbf{s} \cdot \mathbf{H}$, of different signs on the up and down spins.

The energies of the two spin bands are now (see Fig. 5.4)

$$\begin{aligned} \epsilon_{\mathbf{k}\uparrow} &= \epsilon_{\mathbf{k}} + Un_\downarrow + \mu_B H \\ \epsilon_{\mathbf{k}\downarrow} &= \epsilon_{\mathbf{k}} + Un_\uparrow - \mu_B H \end{aligned} \tag{5.18}$$

With the same approximation as before - that the density of states can be taken to be a constant, we can then self-consistently determine the average spin density

$$n_\uparrow - n_\downarrow = [U(n_\uparrow - n_\downarrow) + 2\mu_B H]\frac{1}{2}g(\mu) \quad . \tag{5.19}$$

The magnetisation is $M = \mu_B(n_\uparrow - n_\downarrow)$ which then gives us the static spin susceptibility

$$\chi_\sigma(\mathbf{q} = 0, \omega = 0) = \frac{\mu_B^2 g(\mu)}{1 - \frac{Ug(\mu)}{2}} \quad . \tag{5.20}$$

In the noninteracting case, this is just the Pauli paramagnetic susceptibility of free particles, which is (apart from constants) the same as χ_o (Eq. (4.81)).

In the interacting case, it is enhanced, and will diverge if U is large enough that the Stoner criterion is satisfied

$$\frac{Ug(\mu)}{2} > 1 \quad , \tag{5.21}$$

which marks the onset of ferromagnetism in this model.

5.3.3 Dynamical spin response function in the Hubbard model

The calculations of the uniform susceptibility in the last section can be immediately generalised to calculate the finite $\mathbf{q}, \omega$ response at the same level of approximation. Instead of adding a perturbation term of a static uniform magnetic field, we can imagine adding a dynamic perturbation

$$\mu_B H(\mathbf{q},\omega)(\hat{n}_\uparrow(\mathbf{q},\omega) - \hat{n}_\downarrow(\mathbf{q},\omega)) \quad .$$

If we also make the same mean-field approximation to the Hubbard interaction, then again we have separate Hamiltonians for both up- and down-spins. The up spins see an effective self-consistent potential

$$\mu_B H(\mathbf{q},\omega) + U n_\downarrow(\mathbf{q},\omega) \tag{5.22}$$

and the down spins

$$-\mu_B H(\mathbf{q},\omega) + U n_\uparrow(\mathbf{q},\omega) \tag{5.23}$$

so that the effective field is self-consistently modified by the average spin-fluctuation density through the interaction parameter. We can now get the average spin density from a calculation of the free-electron susceptibility χ_o[4]:

$$n_\uparrow(\mathbf{q}\omega) - n_\downarrow(\mathbf{q}\omega) = [U(n_\uparrow(\mathbf{q}\omega) - n_\downarrow(\mathbf{q}\omega)) + 2\mu_B H]\frac{1}{2}\chi_o(\mathbf{q}\omega) \quad . \tag{5.24}$$

and we thus self-consistently determine the dynamical susceptibility

$$\chi_\sigma(\mathbf{q},\omega) = \frac{\mu_B^2 \chi_o(\mathbf{q},\omega)}{1 - \frac{1}{2}U\chi_o(\mathbf{q},\omega)} \quad . \tag{5.25}$$

The parallel with the last section should be clear; but notice that here we explicitly calculated the response function, rather than evaluating the densities directly from the shifted bands. As $(\mathbf{q}\omega) \to 0$, the results are the same, because we showed earlier that $\chi_o(0,0) = g(\mu)$.

---

[4] Again, the factor of $\frac{1}{2}$ is because χ_o was calculated for both spins

5.3.4 Spin-density waves and antiferromagnets

If we look at the static susceptibility from Eq. (5.25), it should be clear that as we increase U, the first instability that is reached need not be at $q = 0$, but might be at nonzero momentum. In general, the instability will occur with the wavevector corresponding to the maximum value of the static susceptibility $\chi_o(q)$. For free electrons, we saw that three dimensions, the maximum is indeed at $q = 0$, but in one dimension, the peak is at $2k_F$ (actually a singularity). In general the value of the susceptibility will depend on the details of the band-structure, but as a general rule, any quasi-one-dimensionality in the band structure - i.e. strong anisotropy in the dispersion between different directions will lead to peaks at finite momentum.

In a case where the instability occurs at finite q, the ensuing magnetic order will be periodic, and generally this is called a spin-density wave, or an antiferromagnet. (Spin-density wave as a term is usually reserved for cases when the magnetic period is not exactly a lattice vector, and where the amplitude of the magnetic order parameter is small. Examples include a number of quasi-one-dimensional organic metals, and metallic Cr.)

It is worth exploring the one-dimensional case a little further, because it provides a way of connecting to the Mott transition. We saw that in 1D the instability was likely to occur at $2k_F$, and therefore in the spin-density-wave state, there should be a periodic spin density component at a wavevector $2k_F$ (see Fig. 5.5). If there are n electrons per unit cell, note that $k_F = n\pi/2a$, where a is the lattice periodicity. Furthermore, the instability in a 1D model will occur for infinitesimal U, because the response function is singular at $2k_F$.

Now let us imagine increasing U from zero. When it is small, the spin-density-wave state will be of small amplitude: It is best to think of it as two periodic charge density waves, each of period

$$2\pi/2k_F = 2a/n$$

but precisely 180° out of phase. For definiteness, let us take $n = 1$, although the argument will work for other densities too. Because there is a new periodicity in the structure (at $2k_F$), then there is a new Brillouin zone plane (at wavevector k_F), and hence a gap in the single particle bands. The onset of the spin-density-wave is also a metal-insulator transition. Scattering from this periodic structure self-consistently regenerates the spin-density modulation with the correct period. When the amplitude is weak ($U/t \ll 1$), the spin density modulation is very close to being sinusoidal (i.e. a small gap means that the NFE approximation is good); so the charge density remains very nearly uniform. However, if we increase the value of U, we expect the amplitude of the modulation to grow, and therefore the gap to increase.

Eventually the picture (for $U/t \gg 1$ will surely become that shown in the lower panels of Fig. 5.5. The wave can no longer be sinusoidal, because

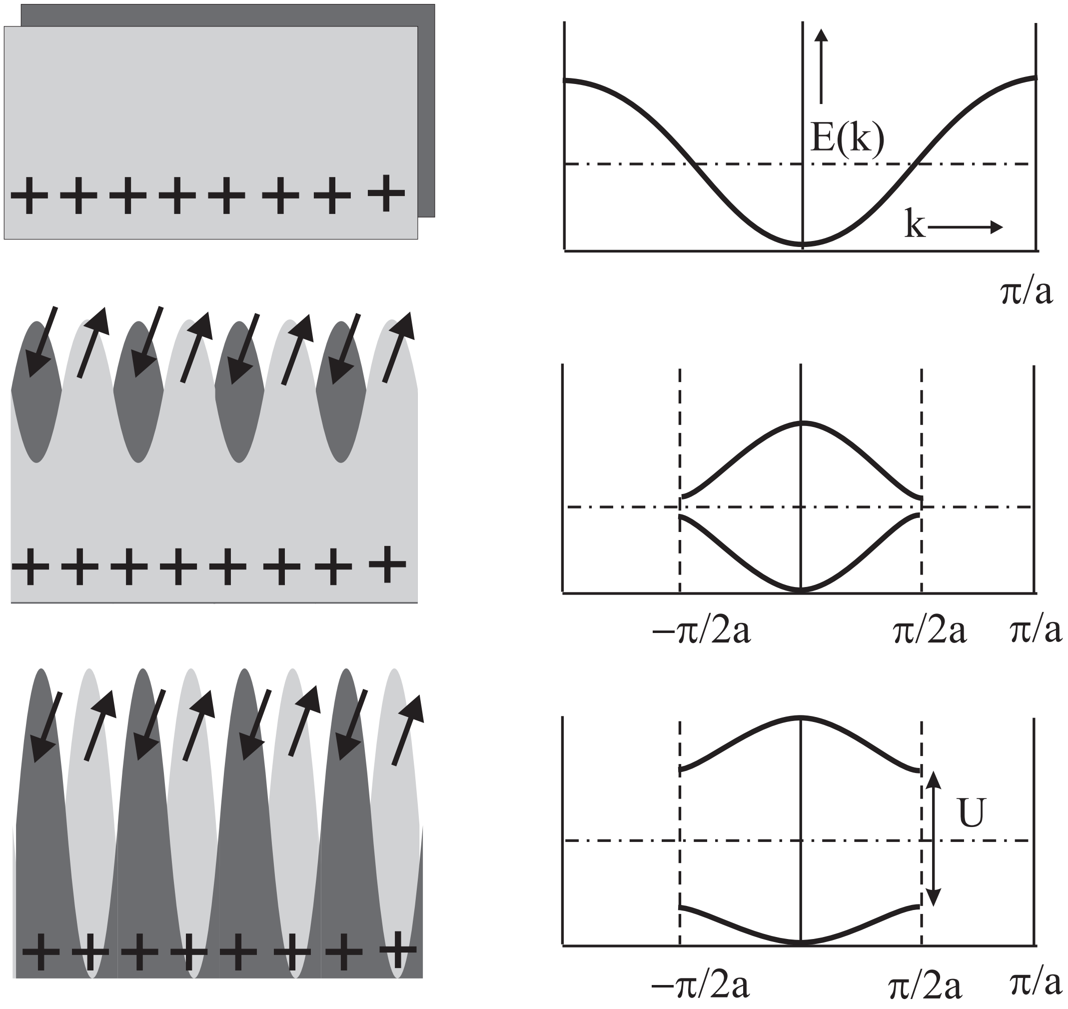


Figure 5.5: One dimensional atomic chain undergoing a spin-density-wave instability. The left panels show charge density for spin up and spin down bands, and the right panels show the bandstructure. In the top panels, we have $U = 0$, and thus a half-filled band (the Fermi energy is marked by the horizontal dash-dot line). In the second panels, $U/t < 1$, and a small gap has opened at the Fermi surface, corresponding to the new magnetic period of $2a$. In the lowest panels, $U/t \gg 1$, and the amplitude of the SDW saturates so that there is exactly one electron on each atomic site; the gap at the Fermi surface is now of order U.

the charge density must remain always positive, for each spin. It will eventually localise to so that each period of the spin density modulation contains precisely one electron, and there is very little overlap from one electron to its neighbour - but of course this is still a ferromagnet. The charge gap is of order U now, because the excitation of a carrier involves moving it from one site onto the neighbouring charge. However, although the ground state is still antiferromagnetic, the (superexchange) interaction which determines the magnetic transition temperature must now be quite small, because it will depend on the overlap of wavefunctions from one electron to its neighbour: it is not difficult to argue that in fact $J_{ij} \approx t^2/U$. This is of course a Mott insulator.

Note the distinction between the two regimes: in 'weak coupling' ($U/t \ll 1$), the instability which produces the antiferromagnetic order also opens a gap at the Fermi surface[5] - but the magnetic interaction is the driving force. In 'strong coupling' ($U/t \gg 1$) it is fundamentally the interaction between charges that produces the Mott transition, and subsequently magnetic order appears on a low energy scale. These two regimes are however smoothly connected.

5.4 Collective magnetic properties

5.4.1 Magnetic phase transitions

The Heisenberg model, however complicated the mechanisms that generate the interactions, provides a very good description of the low energy spin dynamics of many magnetic materials.

For most purposes, and especially to describe phenomena at finite temperatures, it turns out that the spins can be treated classically and so the analysis of magnetic ground states and magnetic ordering becomes a topic in classical statistical physics, that is somewhat removed from the agenda of this course. Because the interaction J is usually small in comparison to other electronic energies in the problem, we need to include the thermal fluctuations only of the spins at low temperatures, because other degrees of freedom are comparatively stiff, so produce only small changes to the free energy at the temperatures where macroscopic magnetic phenomena are seen. The transition temperature of a magnet is determined by a competition between the energetics of the interaction between spins – favouring ordering – and the entropy, which is larger in a disordered state. Only in rare cases do we need to go beyond simple classical models of interacting moments to understand the magnetic behaviour of real materials.

---

[5] Of course, only if the Fermi surface is completely one-dimensional will a small gap lead immediately to an insulator - this is why the antiferromagnetic state of Cr is still metallic

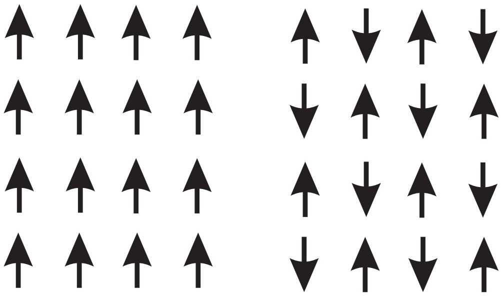


Figure 5.6: . Schematic picture of the ground state of a ferromagnet and an antiferromagnet. The order parameter for the ferromagnet is the uniform magnetisation, and for an antiferromagnet it is the $< \mathbf{S}(\mathbf{Q}) >$, where $\mathbf{Q}$ is the wavevector corresponding to the period of the order

Depending on the sign of J, the ground state will be ferromagnetic (aligned spins) or anti-ferromagnetic (anti-aligned spins on neighbouring sites); more complicated magnetic states can arise if we have different magnetic ions in the unit cell, and also on taking account of magnetic anisotropy.

While it is straightforward to measure the magnetisation in a ferromagnet, measuring the order parameter of an antiferromagnet is more tricky because it corresponds to spins ordering with a finite wavevector. Such order can, however, be cleanly observed by elastic neutron scattering.

5.4.2 Spin waves

The collective excitations are determined by $\chi_\sigma(\mathbf{q},\omega)$ for which Eq. (5.25) is an approximation. It is clear that as one approaches the transition from the paramagnetic side, there must be a build up of oscillator strength at low energies, near the wavevector of the imminent instability. But in general, provided that $q < 2k_F$ all of these features will lie in the continuum, so the spectral function $\Im\chi_\sigma$ will only have peaks which are overdamped.

In the magnetically ordered state, the picture is different, and we will consider here the case of a ferromagnet, starting from the nearest neighbour Heisenberg Hamiltonian

$$H_{Heis} = -J \sum_{i,j=n.n} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j \tag{5.26}$$

The spin variables are angular momenta, and as such they satisfy the commutation relations for angular momenta

$$[\hat{S}_\alpha, \hat{S}_\beta] = i\hbar\epsilon_{\alpha\beta\gamma}\hat{S}_\gamma \quad , \tag{5.27}$$

where the subscripts in Eq. (5.27) refer to the cartesian axes (x, y, z), whereas those in Eq. (5.26) are the site labels.

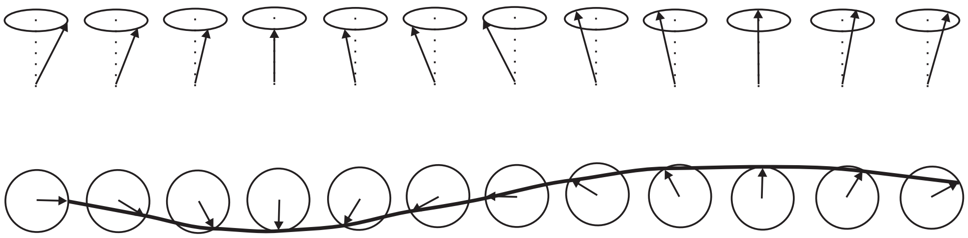


Figure 5.7: The top figure shows the classical picture of a spin wave generated by an array of precessing spins; the bottom figure shows the same viewed from the top.

We can use these two equations to obtain the Heisenberg equation of motion for a spin at a single site

$$\dot{\hat{\mathbf{S}}}_n = -\frac{i}{\hbar}[\hat{\mathbf{S}}_n, H_{Heis}] = \frac{2J}{\hbar} \sum_{j=n.n.\ of\ n} \hat{\mathbf{S}}_n \wedge \hat{\mathbf{S}}_j \tag{5.28}$$

To derive this, we need to assume (correctly) that spin operators on *different* sites commute. The factor of two in this equation comes about because H_{Heis} contains a sum over all pairs i, j, and the commutator will pick out terms with both $i = n$ and $j = n$.

Notice that the form of this equation is that it describes precession of the spin at site n about the effective exchange field $J\sum_j \hat{\mathbf{S}}_j$ of the nearest neighbour spins. While this is a fully quantum mechanical problem, we can gain insight (and the correct answer when the magnitude of the spin is large) by treating this in a semiclassical approximation. [6]

We can assume that in the ordered state there is a uniform ferromagnetic moment $<\mathbf{S}>= S\hat{z}$, which we have chosen to point in the z-direction. We shall now look for the collective modes of small amplitude spin fluctuations about the magnetically ordered ground state. One can guess the form of the solutions by considering a single spin tilted from its axis by a small angle, while the neighbouring spins are held fixed - in this case the spin will simply precess in a circle about the $\hat{z}$-axis. But of course the motion of one spin will affect the others, so the precession of a single spin will not be an eigenstate; but if all of the spins precess in such a way as to keep the relative angles between neighbours constant, then we can have a wave with uniform precessional rate. This is shown pictorially for a one-dimensional chain in Fig. 5.7

To get the dispersion relation for the spin wave mode, we write

$$\hat{\mathbf{S}}_n = S\hat{z} + \delta\mathbf{S}_n \tag{5.29}$$

---

[6] The quantum mechanical case is not much more difficult, but involves making a different representation for the quantisation of the spins; see Marder pp 753-757 .

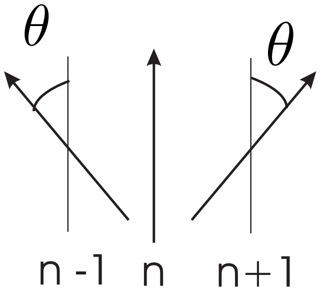


Figure 5.8:

where $\delta\mathbf{S}_n$ is a vector lying in the $x-y$ plane. Substituting this into Eq. (5.28) we get

$$\delta\dot{\mathbf{S}}_n = \frac{2JS}{\hbar} \sum_{j=n.n.\ of\ n} (\delta\mathbf{S}_n - \delta\mathbf{S}_j) \wedge \hat{z} \tag{5.30}$$

The equation in now classical – all the operators have been approximated by classical vectors.

The generalisation to a three-dimensional lattice is quite straightforward:

$$\hbar\omega_{SW}(\mathbf{q}) = 2ZSJ(1 - \frac{1}{Z}\sum_{\mathbf{R}=n.n.} e^{i\mathbf{q}\cdot\mathbf{R}}) \quad , \tag{5.31}$$

where Z is the number of nearest neighbours. Notice that as $\mathbf{q} \to 0$, $\omega \propto q^2$.

We are of course not surprised to find the mode frequency vanishing in the long-wavelength limit, because at $q = 0$, this mode would be a uniform tilting of all the spins in the whole lattice. The Heisenberg model knows only about *relative* directions of spins, so this must have zero energy; our choice of the z-direction for the ordered moment was completely arbitrary. However, the *quadratic* - rather than linear, as for phonons - behaviour is a consequence of a further conservation law – the total spin $\sum_i \hat{\mathbf{S}}_i$ commutes with the Hamiltonian Eq. (5.26). In the case of the Ferromagnet this means the order parameter is *conserved* and the quadratic dependence is a characteristic of ferromagnetic spin waves; spin waves exist also in antiferromagnetically ordered states, but their momentum dependence is indeed *linear* in that case.

5.4.3 Neutron scattering

Neutron scattering is an ideal probe for the observation of typical magnetic fluctuations in solids, because the characteristic energy and momentum range of spin fluctuations is comparable to the energy-momentum dispersion of neutrons available either from reactor or "spallation" sources. Neutrons have a magnetic moment, and therefore interact with magnetic electronic fluctuations, as well as directly with the nucleus.

Reactor sources of neutrons operate continuously, and the energy range of the neutrons is determined by thermalising with a surrounding moderator. This produces beams with a broad band of wavelengths. In a spallation source, neutrons are produced by the bombardment of a heavy metal target with energetic particles from an accelerator. Specific wavelengths can be separated out by scattering from a single crystal monochromator – that operates in essentially the same way as a diffraction grating for light – or by a "chopper" spectrometer that selects the velocity directly.

The neutron-nucleus interaction makes neutron scattering a useful probe for crystal structure determination (elastic) and determining phonon dispersion relations (inelastic). The magnetic interaction allows neutrons to be used as a probe to determine the magnetic ordering of magnetic ground states (by elastic scattering), and to determine the magnetic fluctuation spectrum by inelastic scattering.

Its value is best displayed by showing some data. Fig. 5.9 shows elastic magnetic scattering determining the existence of an antiferromagnetic ordered phase[7]. and Fig. 5.10 shows how inelastic neutron scattering can be used to determine the dispersion relations of spin waves in a ferromagnet[8].

---

[7] R.Plumier, data quoted by Kittel, p 698

[8] T.G. Perring *et al.*, Physical Review Letters **77**, 711 (1996)

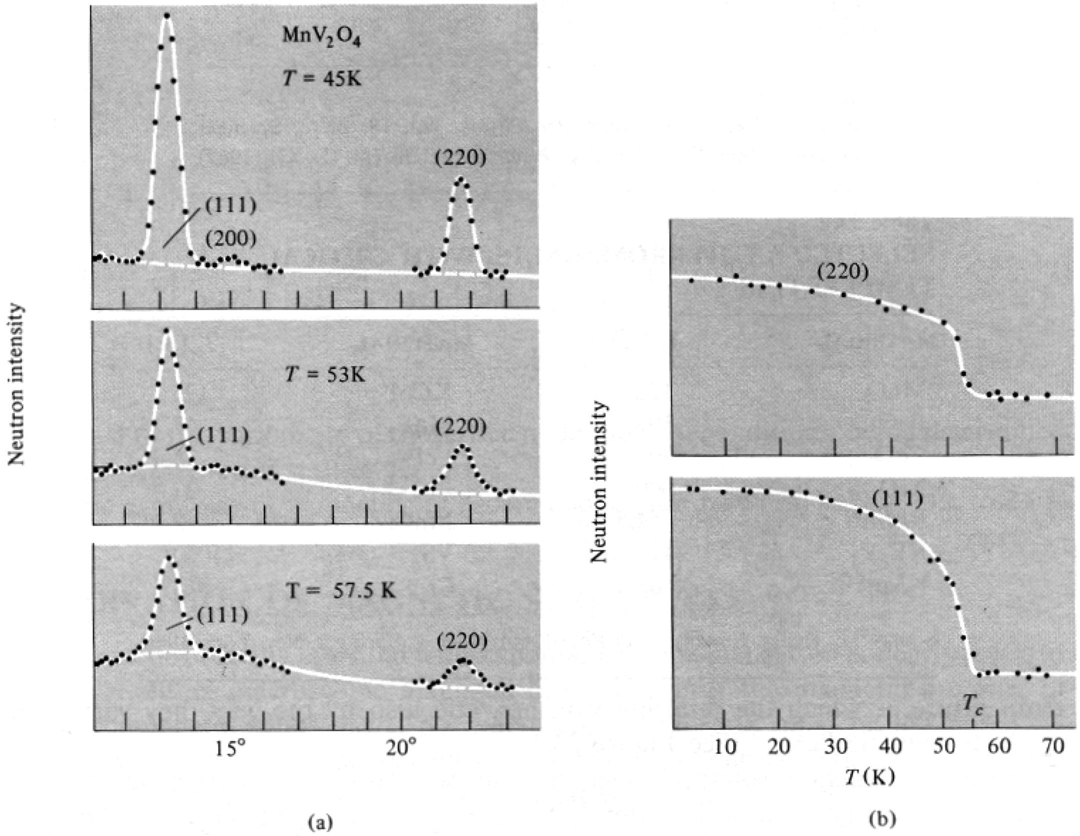


Figure 5.9: Elastic neutron scattering on MnV_2O_4, which is an antiferromagnet with a transition temperature of $T_N = 56K$. The angular scan measures the angle of diffraction of the neutrons, and two magnetic peaks can be seen, that decrease in intensity as the temperature is raised.

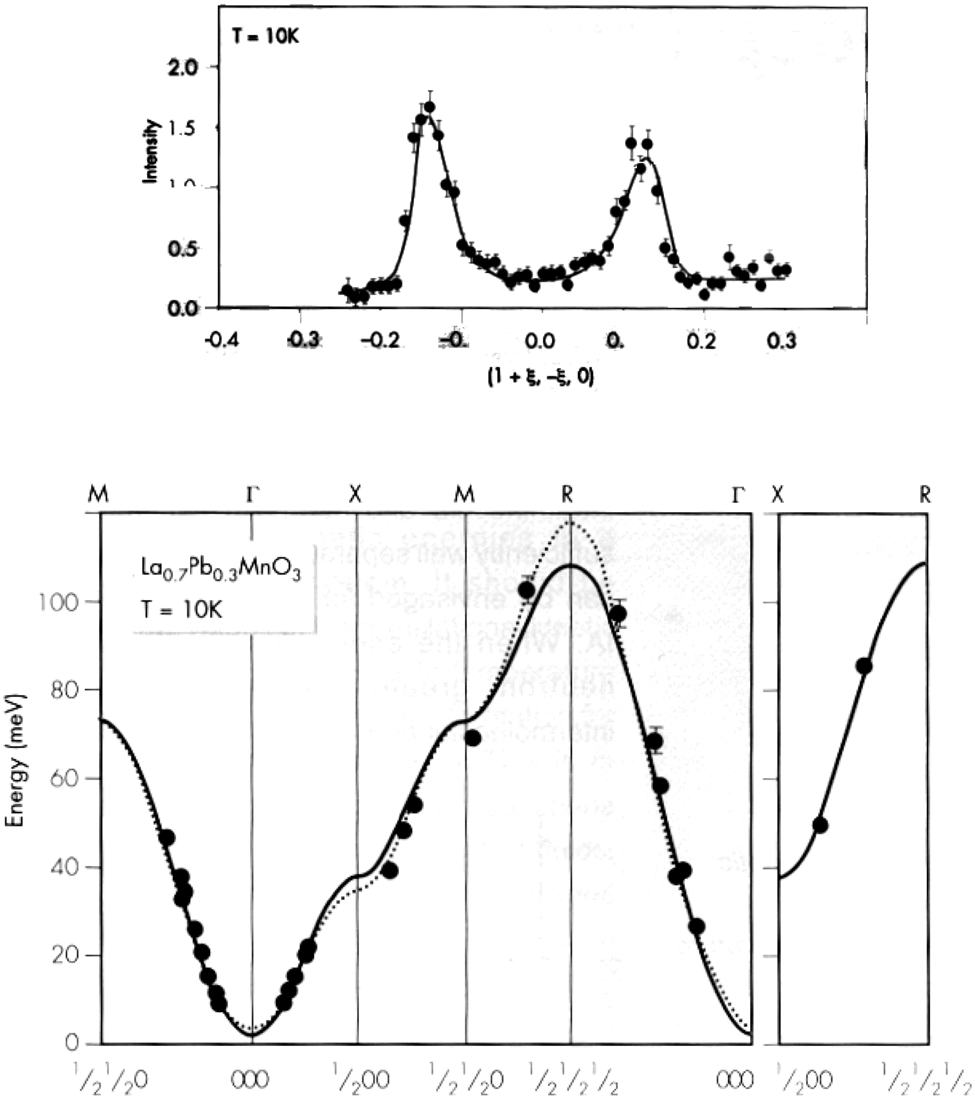


Figure 5.10: Inelastic neutron scattering from the ferromagnet $La_{0.7}Pb_{0.3}MnO_3$, which is well described as a Heisenberg ferromagnet, at 10K. The upper figure shows a slice of the spectrum at a constant energy transfer of approximately 7 meV, as a function of momentum near the (100) reciprocal lattice point. The two peaks correspond to excitation of spin waves of well-defined momentum, with the width of the peaks in the figure given entirely by the experimental resolution. The lower figure maps out the full dispersion relation of the spin waves in the major symmetry directions, using multiple scans of the type shown above (the material is a cubic perovskite, and the magnetic Mn ions lie on an f.c.c. lattice). The solid line is a fit of the dispersion curve to a nearest neighbour Heisenberg model, with an exchange constant $2JS = 9\ meV$; the dotted line corresponds to a slightly different theoretical model – whose distinction from the Heisenberg model is of no consequence to our discussions.

Questions on chapter 5

Qu.5.1 Exchange

Consider single-particle wavefunctions on two neighbouring identical atoms ψ_A, ψ_B, which may be assumed real. These are to be used as the basis for a two-electron state. Show that the charge density in a singlet (triplet) state made out of the two orbitals is given by

$$\rho(r) = |\psi_A(r)|^2 + |\psi_B(r)|^2 \pm 2 < \psi_A|\psi_B > \psi_A(r)\psi_B(r) \quad . \tag{5.32}$$

Explain why the singlet state will usually be lower in energy.

Qu.5.2 One-dimensional spin waves

Assume a one-dimensional chain of spins, precessing according to Eq. (5.30). By considering two neighbours of the n^{th} spin, as in Fig. 5.8, each at relative angles θ, show that the rate of precession according to Eq. (5.30) is

$$\omega = \frac{4JS}{\hbar}(1 - \cos\theta) \quad . \tag{5.33}$$

Hence show that for a spin wave of wavevector q, the dispersion is

$$\hbar\omega = 4JS(1 - \cos(qa)) \quad . \tag{5.34}$$

Qu.5.3 Colossal magnetoresistance

This question introduces an active subject of current research: although lengthy, it involves nothing beyond the material discussed in the lectures. In a material like that shown in Fig. 5.10 the magnetism arises from a mechanism called **double exchange**, which is a version of itinerant exchange but involving two types of d-bands. The prototype compound is $La_{1-x}Sr_xMnO_3$, where the valence of La is 3+ and Sr is 2+. This is a cubic (perovskite) crystal structure where the Mn ions are nominally equidistant from six oxygen neighbours in three cartesian directions.

(a) Explain why the valence of Mn in the compound $La_{1-x}Sr_xMnO_3$ is expected to be between 3+ and 4+ and that the occupancy of the d-levels is expected to be $4 - x$ electrons per Mn ion.

(b) The degeneracy of the 5 d-levels in the free ion is split by the cubic environment into a low energy three-fold degenerate subset (whose notation is t_{2g}) and a higher energy doubly degenerate orbital set (e_g). Explain why the spin configurations of these levels for the Mn^{3+} and Mn^{4+} ions are expected to be as shown in Fig. 5.11.

(c) The lowest three electron states can be regarded as forming a classical spin $S = \frac{3}{2}$ which has negligible hopping from site to site, whereas the highest state is potentially itinerant. Now consider two neighbouring sites i, j in the

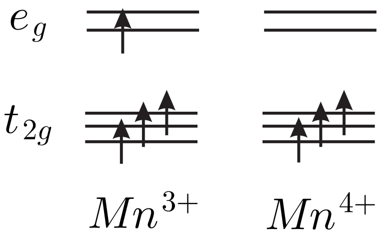


Figure 5.11:

solid, each having the same "core" spin S, and sharing a single itinerant e_g electron, that has a tight-binding matrix element

$$t =< \phi_{e_g}(\mathbf{r}-\mathbf{R}_i)|H|\phi_{e_g}(\mathbf{r}-\mathbf{R}_j) > \tag{5.35}$$

for hopping from site to site.

Explain the origin of the terms

$$H_{int} = -J\sum_i \hat{\mathbf{s}}_i \cdot \mathbf{S}_i + J_x \sum_{ij} \mathbf{S}_i \cdot \mathbf{S_j} \quad , \tag{5.36}$$

in the total Hamiltonian ($\hat{\mathbf{s}}_i$) is the spin of the e_g electron) and suggest relative magnitudes of U, $J$and J_x.[9]

(d) Consider two neighbouring core spins $\mathbf{S}_i$ $\mathbf{S}_j$ that are at a relative angle θ_{ij}. By considering that the spin wavefunction of the itinerant electron must, for $J \gg t$, be always aligned with the local core spin $\mathbf{S}$, explain why the Schrödinger equation for the itinerant electron can be simplified to one in which the tight-binding hopping matrix element from site i to site j is replaced by

$$t_{eff} = t\cos(\frac{\theta_{ij}}{2}) \quad . \tag{5.37}$$

To do this, you may wish to note that under a rotation by an angle θ, the spin wavefunction transforms as

$$\begin{pmatrix} |\uparrow'> \\ |\downarrow'> \end{pmatrix} = \begin{pmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} |\uparrow> \\ |\downarrow> \end{pmatrix} \tag{5.38}$$

(e) Sketch the density of states of the itinerant electrons for different alignments of the core spins $\mathbf{S}$:
ferromagnetic (all core spins aligned),
antiferromagnetic (all neighbouring core spins anti-aligned),

---

[9] In second-quantised notation, the full Hamiltonian can be written as

$$H = t\sum_{ij=n.n.,\sigma} \hat{c}^{\dagger}_{i\sigma}\hat{c}_{j\sigma} + U\sum_i \hat{n}_{i\sigma}\hat{n}_{i-\sigma} - J\sum_i \hat{\mathbf{s}}_i\cdot\mathbf{S}_i + J_x\sum_{ij}\mathbf{S}_i\cdot\mathbf{S_j} \quad .$$

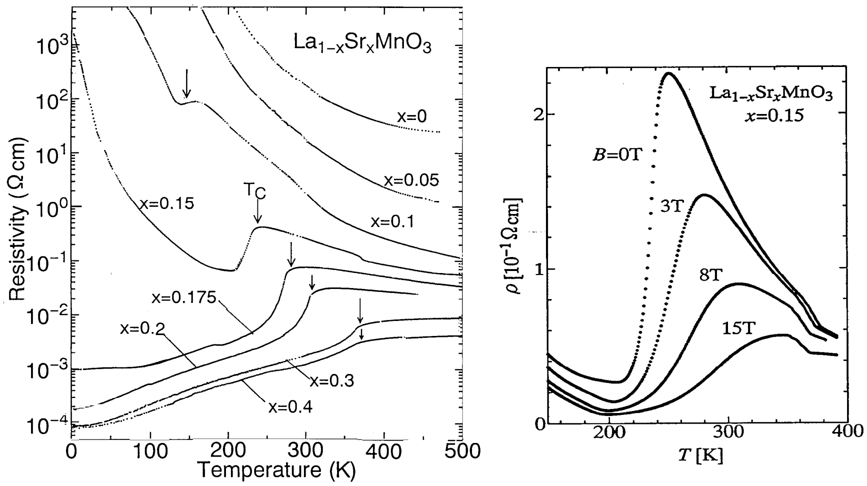


Figure 5.12: Resistivity as a function of temperature and magnetic field for a series of samples of doped manganese oxides with different compositions. The ferromagnetic transition temperatures T_c are marked by the arrows.

paramagnetic (core spins randomly aligned).
Discuss how the total Free energies of these states differ, and suggest what is the magnetic ground state when $x = 0$; and when $tx > J_x$; give rough estimates of the transition temperatures of the ordered magnetic states toward high temperature paramagnetism.

(f) Fig. 5.12 shows the resistivity as a function of temperature of several samples of $La_{1-x}Sr_xMnO_3$ with different concentrations x, as well as the magnetic-field-dependence of the resistivity (which gives rise to the label "colossal" magnetoresistance).[10] Discuss this data in light of the results above.

---

[10] Urushibara *et al.* Physical Review B **51** 14103 (1995)

Chapter 6

Electrons and Phonons

6.1 Electron-phonon interaction

The movement of atoms from their equilibrium position generates an electrostatic potential from which the electrons will scatter, and this interaction is generically called the electron-phonon interaction. This is a large subject, and we will cover only the very basics in this course.

We will begin with a somewhat heuristic[1] derivation of the electron-phonon interaction.

Without being specific about the details, it should be clear that the induced potential due to a lattice displacement $u(\mathbf{r}) = u_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}}$ will be of the form

$$V(\mathbf{q},\omega) = g_{\mathbf{q}} u_{\mathbf{q}\omega} \quad , \tag{6.1}$$

where $g_{\mathbf{q}}$ is the *electron-phonon coupling constant*, a function that can in principle be calculated using the tools of Section 2. For later reference, we should note that this potential should be regarded as *screened* by the other electrons:

$$V = V_{ext}/\epsilon(\mathbf{q}\omega) \tag{6.2}$$

The Hamiltonian for electrons interacting with the phonons should now be familiar:

$$H_{e-ph} = H_o - \sum_{\mathbf{q}\omega} V(\mathbf{q},\omega)\hat{\rho}_{\mathbf{q}\omega} \quad . \tag{6.3}$$

where H_o contains all of the terms involving interactions of the electrons with the static lattice potential, and with each other. The phonons appear as a time dependent potential coupling to the electronic system.

In the electron-phonon interaction, one should remember that not only do the phonons influence the electrons, but the electrons influence the phonons,

---

[1] I will treat the lattice as a classical medium, whereas a more correct derivation requires second quantisation of the phonons.

because the last term in Eq. (6.3) will induce a force on the atoms proportional to the electron density perturbation. In order to describe the phonons, we will simply need to know the response function to an external density perturbation, i.e a phonon susceptibility which will be of the general form [2]

$$\chi_{ph}(\mathbf{q},\omega) = \frac{1}{-\omega^2 + \omega(\mathbf{q})^2} \quad , \tag{6.4}$$

where $\omega(\mathbf{q})$ describes the dispersion of the phonon mode of wavevector $\mathbf{q}$.

6.2 Effective interaction between electrons

One of the most important features of the electron-phonon interaction is that it generates an effective interaction between electrons.

This can be visualised as follows: A quantum density fluctuation in the electron gas will couple to a local lattice distortion by the second term in Eq. (6.3); the magnitude and time dependence of this distortion will be governed by the interaction strength $g_{\mathbf{q}}$, and resisted by the stiffness of the lattice χ_{ph}^{-1} (see Fig. 6.1).

This lattice distortion, once formed will last for a characteristic time scale $\approx \omega_{phonon}^{-1}$, but the density fluctuation in the electron gas has a much shorter lifetime $\approx E_F^{-1}$. However, the very presence of the long-lived lattice distortion will of its own account encourage further quantum fluctuations of the electron density – this means that the interaction with the lattice produces correlations of the density fluctuations in the electron gas, just as if there were an *attractive* potential between electrons.

We can work this out straightforwardly from the Hamiltonian Eq. (6.3) and the phonon response function.

First, we calculate the lattice displacement induced by a density fluctuation in the electron gas:

$$u_{\mathbf{q}\omega} = g_{\mathbf{q}} \chi_{ph}(\mathbf{q}\omega) \hat{\rho}_{\mathbf{q}\omega} \tag{6.5}$$

Then we may substitute for u in Eq. (6.3) , and get an effective Hamiltonian which no longer depends on the lattice:

$$H_{eff} = H_o - \sum_q |g_{\mathbf{q}}|^2 \chi_{ph}(\mathbf{q}\omega) \hat{\rho}_{\mathbf{q}\omega} \hat{\rho}_{-\mathbf{q}-\omega} \quad . \tag{6.6}$$

Eq. (6.6) has now produced a *dynamical* interaction between the electrons which should be added to the Coulomb interaction (which is contained

---

[2] It is conventional to define χ_{ph} with a factor of $\omega(\mathbf{q})$ in the numerator. However, for our purposes this will be swept up into the coupling constant later on.

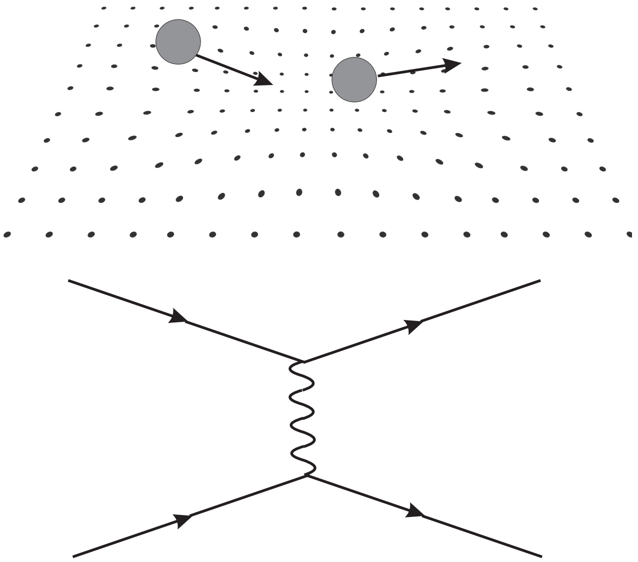


Figure 6.1: Schematic picture of the electron-phonon interaction, and how it induces a retarded, attractive force between electrons. The top figure illustrates the distortion induced by a charge fluctuation, which persists long enough to attract another electron into its wake. The bottom figure is the diagrammatic description.

in H_o). Going back to real space, the total (electron-electron plus phonon-mediated) interaction between two particles is just

$$V_{e-e+e-ph-e}(\mathbf{r}_1,\mathbf{r}_2) = \frac{e^2}{|\mathbf{r}_1-\mathbf{r}_2|} - \sum_{\mathbf{q}} e^{i\mathbf{q}\cdot(\mathbf{r}_1-\mathbf{r}_2)}|g_{\mathbf{q}}|^2\chi_{ph}(\mathbf{q}\omega) \quad . \tag{6.7}$$

however, it depends on frequency. If the electrons are in well-defined quasiparticle states, of momenta $\mathbf{k}$, $\mathbf{k}'$, then the relevant energy is $\omega = \epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}'}$ – this is most clearly seen in the schematic Fig. 6.1, where the interaction transfers energy $\hbar\omega$ and momentum $\mathbf{q}$ from one particle to the other.

The total interaction is sketched in Fig. 6.2 as a function of frequency. Note that the effect of the phonons is to reduce the interaction between electrons on energy scales just below the phonon frequency, and for a finite range of energy, to make the overall interaction *attractive*. This overscreening is of crucial importance in generating the attractive interaction between electrons that leads to superconductivity.

Notice that the interaction only looks attractive because it is *retarded*. The instantaneous (i.e. $\omega \to \infty$) interaction between electrons is still the repulsive Coulomb interaction. It is very important here that $\omega_{ph} \ll E_F$, because the interaction between electrons is then between a charge density fluctuations at *different* times, and can be attractive for times greater then

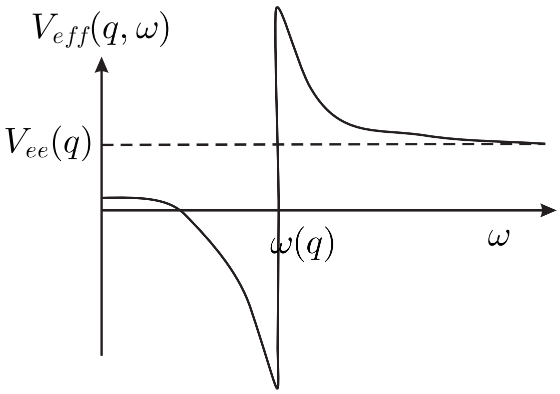


Figure 6.2: Dynamical electron-electron interaction

ω_{ph}^{-1}.

6.3 Effective mass enhancement

The interaction between particles in Eq. (6.6) is of the form

$$v_{eff}(\mathbf{k},\mathbf{k}') = -|g_{\mathbf{q}}|^2\chi_{ph}(\mathbf{q},\omega), \qquad \mathbf{q}=\mathbf{k}-\mathbf{k}', \quad \omega = \epsilon_{\mathbf{k}} - \epsilon_{k'} \quad . \tag{6.8}$$

Notice that the interaction couples quasiparticles with *different* energies, and furthermore that it is operative only for those states within a narrow shell of width $\hbar\omega_D$ (ω_D is the Debye frequency) of each other.

This interaction gives rise to a renormalisation of the quasiparticle energy, and the simplest approximation is to use the Hartree-Fock approximation that we used earlier for the effect of the Coulomb interaction. This will give an energy shift (see Eq. (2.102))

$$\Delta\epsilon_{\mathbf{k}} = -\int \frac{d\mathbf{k}'}{(2\pi)^3} n_{\mathbf{k}'} v_{eff}(\mathbf{k},\mathbf{k}') \quad , \tag{6.9}$$

which is (μ is the chemical potential, as usual)

$$\epsilon_{\mathbf{k}} - \mu = {\epsilon_{\mathbf{k}}}^o - \mu - \int \frac{d\mathbf{k}'}{(2\pi)^3} \frac{|g_{\mathbf{k}-\mathbf{k}'}|^2 n_{\mathbf{k}'}}{(\epsilon_{\mathbf{k}}-\epsilon_{\mathbf{k}'})^2 - \omega(\mathbf{k}-\mathbf{k}')^2} \quad , \tag{6.10}$$

where ${\epsilon_{\mathbf{k}}}^o$ is the quasiparticle dispersion in the absence of the electron-phonon interaction.

Details of the algebra are relegated to a question, but one can estimate the effects straightforwardly. It is important to remember that usually the electronic bandwidth is much larger than any characteristic phonon frequency ω_D (say the Debye frequency for acoustic phonons). Then it should be clear that the effects of the electron phonon interaction are small for states far from the chemical potential. If $\epsilon_{\mathbf{k}} - \mu \gg \omega_D$, then since the effective potential has to be generated by *occupied* states – note the factor of $n_{\mathbf{k}'}$ in Eq. (6.10) – the energy shift gets small. It is actually of order

$$\frac{g^2 N(\mu)}{\epsilon_{\mathbf{k}} - \mu} \tag{6.11}$$

where $N(\mu)$ is the density of states in energy[3].

When the energy is within the Debye frequency of the chemical potential, the effects can be large. For $|\epsilon_{\mathbf{k}} - \mu| \ll \omega_D$,

$$\epsilon_{\mathbf{k}} - \mu = \frac{{\epsilon_{\mathbf{k}}}^o - \mu}{1 + \lambda} \tag{6.12}$$

where

$$\lambda = \frac{g^2 N(\mu)}{\omega_D^2} \tag{6.13}$$

is a dimensionless coupling constant. The velocity of the electronic dispersion in the vicinity of the chemical potential is reduced – as if the particles became heavier. In fact they do, because the physics is that they are moving surrounded by a cloud of lattice polarisation – but quasiparticles that are moving at energies much more than the phonon frequency above the chemical potential escape from the cloud, and are not renormalised by it.

For most materials, the dimensionless electron-phonon coupling constant λ is less than or of order unity, but nevertheless gives a substantial renormalisation of the energy spectrum near the Fermi surface, but which rapidly vanishes at higher energies. The characteristic correction is shown in Fig. 6.3.

In the vicinity of the Fermi surface, the carrier velocity is reduced, and hence there is an enhancement of the density of states

$$N^*(\mu) = N(\mu)(1 + \lambda) \quad , \tag{6.14}$$

or equivalently of the quasiparticle effective mass

$$m^* = m(1 + \lambda) \quad . \tag{6.15}$$

These corrections apply only to energies within ω_D of the Fermi energy. The enhanced effective mass feeds into other thermodynamic properties, although one cannot apply the correction indiscriminately. It turns out that the specific heat

$$C_{el}/T = \frac{\pi^2}{3} k_B^2 N^*(\mu) \tag{6.16}$$

is renormalised, but the Pauli susceptibility

$$\chi_{Pauli} = \mu_B^2 N(\mu) \tag{6.17}$$

is not enhanced.

---

[3] We change notation from the $g(\epsilon)$ we have used before to avoid confusion with the coupling constant $g_{\mathbf{q}}$

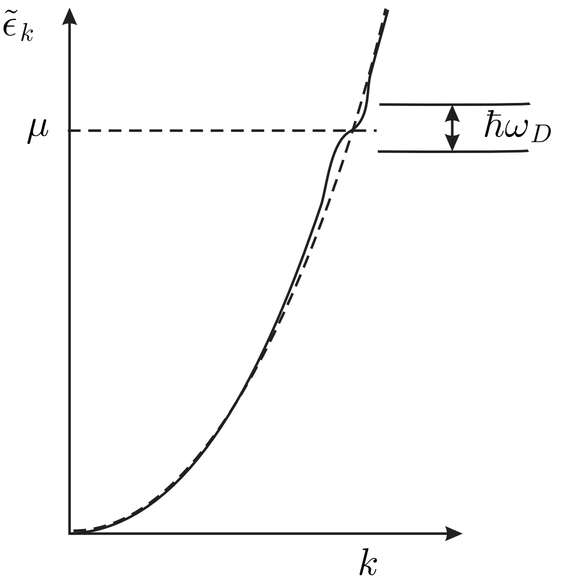


Figure 6.3: Phonon renormalisation of the electronic dispersion

6.4 Cooper's problem

If electrons close to the Fermi surface can experience an effective attraction this leads to the idea that two electrons might bind to form a pair, and this was shown by Cooper[4]. In fact the idea that even a very weak attraction might produce a bound pair seems not reasonable, because we know that (in three dimensions) a two body bound state does not form unless the attractive potential exceeds a certain strength. It turns out that one gets a bound state for an arbitrarily weak interaction, *because of the presence of the Fermi sea.* We will go through the algebra, but the answer can be found at the end of this section.

The general wavefunction for a bound pair can be written

$$\psi_C(\mathbf{r}_1,\mathbf{r}_2) = \sum_{\mathbf{k}_1\mathbf{k}_2} g(\mathbf{k}_1,\mathbf{k}_2) e^{i\mathbf{k}_1\cdot\mathbf{r}_1} e^{i\mathbf{k}_2\cdot\mathbf{r}_2} \tag{6.18}$$

but if we restrict ourselves to states with zero centre-of-mass momentum, we shall choose $\mathbf{k}_2 = -\mathbf{k}_1$, and write $g(\mathbf{k},-\mathbf{k}) = g(\mathbf{k})$. Furthermore, if we are looking for a bound state, we expect to put the pair into a singlet state, so that the spatial part of the wavefunction is nodeless – and the electrons in the pair have a greater probability of being close to one another.

Cooper's problem includes an attractive interaction between two electrons added to the Fermi sea, but neglects any interactions of those states with the Fermi sea itself. The only use of the Fermi sea is to prevent the

---

[4] L.Cooper, Physical Review **104**, 1189 (1956)

added electrons occupying states inside the Fermi surface. Of course this is not strictly legal, but nevertheless it is very helpful as a toy problem.

In the momentum representation, the two particle state can be an eigenstate of energy E if

$$(E - 2\epsilon_k)g_k = \sum_{k' > k_F} V_{k-k'} g_{k'} \quad . \tag{6.19}$$

This equation can be solved with the simplifying approximation of taking $V_{k-k'} = -V$, a negative constant for k-states within ω_D of μ. Then the RHS of Eq. (6.19) is just a constant, and

$$g_k = V \frac{\sum' g_{k'}}{2\epsilon_k - E} \tag{6.20}$$

where the prime on the summation means the restriction $\mu + \omega_D > \epsilon_{k'} > \mu$. Summing this equation, we can then cancel the $\sum g_{k'}$ and get

$$\frac{1}{V} = \sum{}' \frac{1}{2\epsilon_k - E} \quad . \tag{6.21}$$

We can replace the sum over momentum by an integral over the energy, with $N(0)$ the density of states at the fermi energy:

$$\frac{1}{V} = N(0) \int_{\mu}^{\mu+\omega_D} \frac{d\epsilon}{2\epsilon - E} = \frac{1}{2} N(0) \ln \left| \frac{2\mu - E + 2\omega_D}{2\mu - E} \right| \quad . \tag{6.22}$$

Finally if we take the weak coupling limit $N(0)V \ll 1$, we get

$$E = 2\mu - 2\omega_D e^{-2/N(0)V} \quad . \tag{6.23}$$

Despite having used only electrons whose kinetic energy was greater than μ, we have a bound state just below the fermi surface. Though weak, the binding energy is finite even for an infinitesimally small attractive interaction. And lastly, the binding energy is a non-analytic function of V as $V \to 0$; this result cannot be obtained to any finite order in perturbation theory.

Cooper's result was the breakthrough that led rapidly to the BCS[5] theory of superconductivity. The fundamental aspects of the Cooper problem – bound pairs, therefore a gap – persist in the BCS theory, but the BCS ground state goes further in having a collective broken symmetry.

---

[5] J.Bardeen, L.N.Cooper and J.R.Schrieffer, Physical Review **108**, 1175 (1957)

Questions on chapter 6

Qu.6.1 Electron-phonon interaction

Write short notes explaining the physical effects that may be produced by the electron-phonon interaction in metals.

Qu.6.2 Electronic mass enhancement

The integral in Eq. (6.10) can be approximated by neglecting the momentum dependence of the coupling constant g, and replacing the phonon frequency by the characteristic scale ω_D. Show that in this case the integral becomes

$$g^2 \int_{-\infty}^{\mu} d\epsilon' N(\epsilon') \frac{1}{(\epsilon' - \epsilon_{\mathbf{k}})^2 - \omega_D^2} \quad . \tag{6.24}$$

where $N(\epsilon)$ is the density of states in energy[6] . Since the dominant part of the integral comes from energies near the Fermi energy, we can usually replace $N(\epsilon)$ by $N(\mu)$. Making this approximation, show that

(a) For energies $|\epsilon_{\mathbf{k}} - \mu| \ll \omega_D$,

$$\epsilon_{\mathbf{k}} - \mu = \frac{\epsilon_{\mathbf{k}}{}^{o} - \mu}{1 + \lambda} \tag{6.25}$$

where

$$\lambda = \frac{g^2 N(\mu)}{\omega_D^2} \quad . \tag{6.26}$$

(b) For energies $|\epsilon_{\mathbf{k}} - \mu|$ several times ω_D the correction to $\epsilon_{\mathbf{k}}$ is of order

$$\lambda \frac{\omega_D^2}{(\epsilon_{\mathbf{k}} - \mu)^2} (\epsilon_{\mathbf{k}} - \mu) \tag{6.27}$$

Qu.6.3 * Cooper's problem

For those who want to work out a non-trivial problem using operator techniques, this is Cooper's problem done that way. It is good practice of your operator technique to go through the following manipulations of $H|\psi_C>$, although this is pretty hard. But if you follow the rules, it all works out in about a page of algebra.

The wavefunction of a Cooper pair of electrons added to the Fermi sea is

$$|\psi_C> = \sum_{\mathbf{k}>k_F} g_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k}\uparrow} \hat{c}^{\dagger}_{-\mathbf{k}\downarrow} |FS> \quad , \tag{6.28}$$

where only terms in the sum for $k > k_F$ are allowed.

---

[6] We change notation from the $g(\epsilon)$ we have used before to avoid confusion with the coupling constant $g_{\mathbf{q}}$

We can now test out the pair wavefunction with the Hamiltonian

$$H = \sum_p \epsilon_p \hat{c}^\dagger_p \hat{c}_p + \frac{1}{2} \sum_{pp'q} V_q \hat{c}^\dagger_p \hat{c}^\dagger_{p'} \hat{c}_{p'-q} \hat{c}_{p+q} \quad . \tag{6.29}$$

applied to the two electrons in question, but leaving the fermi sea inert. V_q is here taken to be an *attractive* interaction.

Show that the first term in Eq. (6.29) operating on $|\psi_C>$ is

$$\begin{aligned} H_o|\psi_C> &= \sum_{pk\sigma} \epsilon_p g_k \hat{c}^\dagger_{p\sigma} \hat{c}_{p\sigma} \hat{c}^\dagger_{k\uparrow} \hat{c}^\dagger_{-k\downarrow} |FS> \\ &= \sum_k 2\epsilon_k g_k \hat{c}^\dagger_{\mathbf{k}\uparrow} \hat{c}^\dagger_{-\mathbf{k}\downarrow} |FS> \quad . \end{aligned} \tag{6.30}$$

(Hint: the trick in all of these operator manipulations is to move the annahilation operator to the RHS, so that it can destroy the vacuum state. Along the way, it has to anticommute with the creation operators initially on its right and these anticommutators always generate an extra δ-function. The two terms in the last equation come because we must have either $p = k$, $\sigma = \uparrow$, or $p = -k$, $\sigma = \downarrow -$ and $\epsilon_{-p} = \epsilon_p$. Remember that for this toy problem **alone**, we don't apply the Hamiltonian to the Fermi sea.)

Similarly, show that the operation of the second term in Eq. (6.29) gives

$$H_{int}|\psi_C> = \sum_{kpp'q\sigma\sigma'} V_q g_k \hat{c}^\dagger_{p\sigma} \hat{c}^\dagger_{p'\sigma'} \delta_{p+q,k} \delta_{\sigma\uparrow} \delta_{p'-q,-k} \delta_{\sigma'\downarrow} |FS> \tag{6.31}$$

$$= \sum_{kk'>k_F} V_{k-k'} g_{k'} \hat{c}^\dagger_{k\uparrow} \hat{c}^\dagger_{-k\downarrow} |FS> \quad . \tag{6.32}$$

Getting to the final equation involves a little crafty relabelling of the momenta in the sum.

This gets us to the two-particle Schrödinger equation Eq. (6.19).

Chapter 7

Recapitulation

These are some of the things that you should have seen running through the course. The order that these are written down is not the same as you saw in the presentation of the material in the course - but you should be able to follow the connections. Some topics belong in several places and thus appear more than once. This is anyway a partial list, but you may find it useful to organise your thoughts.

For the purposes of revision, you should consider that you may need to explain *everything*, but complicated derivations and proofs (e.g. Bloch's theorem, Hartree-Fock equations (in first quantised form), proof of density functional theory, calculation of the general form of the linear response function for electron gas, fluctuation-dissipation theorem, proof of Kramers-Kronig relations, electron-phonon mass enhancement and Cooper pair) will not be required for examination. *Sketches* of the derivations that explain the principles involved *may* be required.

7.1 Response functions

1. Definition
$$u(\mathbf{q},\omega) = \chi(\mathbf{q},\omega)F(\mathbf{q},\omega)$$
2. Examples for a classical field, e.g. acoustic phonons, plasmons, optical conductivity, reflectivity of metals, spin waves, Thomas-Fermi screening.
3. The origin of the response function in a classical equation of motion.
4. What the response function contains
    - (a) Collective modes
    - (b) Absorption of energy, inelastic scattering
    - (c) Stability criterion at $\omega = 0$

   (d) Relationship to correlation function and fluctuation-dissipation theorem
   (e) Causality and Kramers-Krönig relations.

5. Density response function of the electron gas
   (a) Motivate the formula (for the free electron gas)

$$\chi_o(\mathbf{q},\omega) = 2\sum_{\mathbf{k}} \frac{(n_{\mathbf{k}} - n_{\mathbf{k+q}})}{(\epsilon_{\mathbf{k+q}} - \epsilon_{\mathbf{k}}) - \omega - i\eta}$$

   Behaviour of $\chi_(q,0)$ in 1,2,3 dimensions.
   (b) Use of χ_o in self-consistent field theories
      - For the charge (dielectric response)

$$\chi_\rho = \chi_o/(1 + v_{ee}(q)\chi_o)$$

      Hence $\epsilon^{RPA} = 1 + v_{ee}(q)\chi_o$; limits at long wavelengths and low frequencies

$$\epsilon(q,0) = 1 + q_{TF}^2/q^2$$
$$\epsilon(0,\omega) = 1 - \omega_p^2/\omega^2$$

      and simple derivations of these results. Friedel oscillations $2k_F$, particle-hole continuum excitations
      - Spin response In the Hubbard model (short range interactions).

$$\chi_\sigma = \chi_o/(1 - U\chi_o/2)$$

         - Relation to Hartree-Fock, static calculation of Pauli paramagnetic susceptibility.
         - Instabilities - Stoner Ferromagnetism, Spin Density Wave, antiferromagnetism, connection to Mott insulator
         - Spin waves in the ferromagnet.
      - Density waves/Peierls

$$\chi_\rho = \chi_o/(1 + U_{eff}\chi_o)$$

      Note that U_{eff} negative due to electron-phonon interaction:

$$U_{eff} = v_{ee}(q) - g_q^2\chi_{ph}(q,\omega)$$

   (c) Electron-phonon interaction
      - Dynamical effective interaction
      - Mass enhancement
      - Cooper pair
      - Peierls instability, looked at from several points of view
         - Direct calculation of the energy from the bandstructure with an added potential from the lattice displacement
         - From the static or dynamic susceptibility

7.2 Fermi liquids

1. Electronic band structure of independent particles.
    - Bloch's theorem and the existence of bands, with $E(k)$ a continuous function of k.
    - Density of states in k and in energy.
    - Fermi surface.
    - Interactions in the mean-field Hartree- and Hartree-Fock approximation.
    - Sketch of density functional theory, and the meaning (or not) of energy bands in this case.
2. Quasiparticles.
    - Definition of electron spectral function $A(p, \omega)$.
    - Renormalisation of the energy bands for independent particles; mass, lifetime, spectral weight. Effects of the above on specific heat, paramagnetic susceptibility.
    - Photoemission.

7.3 Metal-Insulator transitions, local moments, and magnetism

1. Jellium. Instability to ferromagnetic metal (in HF), and to Wigner crystal. Stoner-Hubbard model of ferromagnetism. (same as Hartree-Fock, but with a screened short-range Coulomb interaction, U)
2. Screening of the Coulomb interaction: Thomas-Fermi in electron gas; on-atom screening of effective U and the Hubbard Hamiltonian.
3. Mott insulator in the Hubbard model.
    - Simple argument comparing energy of localised versus delocalised electrons at half-filling.
    - Connection of the Mott insulator to the spin-density-wave instability of the electron gas.
    - Applicability of the model, origin of U (screening), value of U/t etc.
4. Local moment magnetism
    - Origin of local magnetic moments - direct exchange and Hund's rules.

  - Interactions between moments - direct exchange, superexchange, itinerant exchange.

5. Collective modes in ferromagnets - spin waves.
6. Elastic and inelastic neutron scattering as probe of magnetic order and fluctuations.

7.4 Materials

1. Basic principles of the bandstructure of simple metals, d-band metals, and semiconductors.
   - Electron counting, band gaps, etc.
   - Bonding and antibonding states in simple models; covalent and ionic bonds.
2. Interaction parameters and screening for transition metals and transition metal oxides

Appendix A

Mathematical appendices

These are a brief recapitulation of the principles involved, together with some definitions and generalisation to three dimensions. If you are uncomfortable with the details, look back at your 1B lecture notes (either in Maths for Nat Sci, or the ?? course)

Other useful sources are the examples sheets of the 1B mathematical physics examples classes (see especially II & III for Fourier series and transforms, VII for variational methods, and IX for applications of quantum mechanics).

A.1 Fourier series and Fourier transforms

A.1.1 Fourier's Theorem

Take a function $f(x)$ that is periodic with period a:

$$f(x+a) = f(x)\ .$$

This can be expanded in terms of any complete set of orthogonal functions with the same period. One choice is the set of complex exponentials, in terms of which *Fourier's theorem* (in its complex form) states that:

$$f(x) = \sum_{-\infty}^{\infty} b_n e^{i2\pi nx/a} \tag{A.1}$$

$$b_n = \frac{1}{a}\int_{-a/2}^{a/2} dx f(x)\, e^{-i2\pi nx/a}\ . \tag{A.2}$$

In crystals, the (one-dimensional) set of points $\{G_n = 2\pi n/a\}$ form a one-dimensional set that makes up the reciprocal lattice. Also we often indulge in the (bad) habit of using the same label for the function as its transform, distinguishing the function and its transform by its argument.

Consequently, a one-dimensional periodic potential $V(x) = V(x+a)$ will be written as

$$V(x) = \sum_{G=-\infty}^{\infty} V(G) e^{iGx} \tag{A.3}$$

$$V(G) = \frac{1}{a}\int_{-a/2}^{a/2} dx\, V(x)\, e^{-iGx} \tag{A.4}$$

Generalising this to a three-dimensionally periodic structure with primitive lattice vectors $\mathbf{a}_i$ gives

$$V(\mathbf{r}) = \sum_{\mathbf{G}} V(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \tag{A.5}$$

$$V(\mathbf{G}) = \frac{1}{V_{cell}}\int_{\text{cell volume}} d\mathbf{r}\, V(\mathbf{r})\, e^{-i\mathbf{G}\cdot\mathbf{r}} \tag{A.6}$$

where the $\mathbf{G}_i$ lie on the three-dimensional reciprocal lattice formed from the primitive vectors

$$\mathbf{b}_1 = 2\pi \frac{\mathbf{a_2} \wedge \mathbf{a_3}}{\mathbf{a}_1 \cdot \mathbf{a}_2 \wedge \mathbf{a}_3} \quad \text{and cyclic permutations} \ . \tag{A.7}$$

This should be obvious if the lattice vectors are orthogonal; it is true in general, and it is easiest to see why by considering the Fourier transform below.

A.1.2 Fourier transforms

A Fourier series arises when you consider a periodic function; it can be made up only of sine/cosine waves of periods that are divisors of the basic period (or wavevectors that are multiples). Bit if you have a function that is not periodic, you need to include waves of all possible wavelengths to make up the function. Eq. (A.1) and Eq. (A.2) can also be used to develop a Fourier transform of a function extending through the whole body of the solid. In one-dimension, for a system of length L, we periodically replicate that large system by applying periodic boundary conditions, viz

$$f(x+L) = f(x) \tag{A.8}$$

(remember that L will be a very long length compared to the lattice constant a).

Rewrite Eq. (A.1) and Eq. (A.2) by defining

$$\tilde{f}(k_n) = Lb_n \ \text{ where } \ k_n = 2\pi n/L \ \ , \tag{A.9}$$

so that

$$f(x) = \frac{1}{2\pi}\sum_n \frac{2\pi}{L}\tilde{f}(k_n)e^{ik_nx} \tag{A.10}$$

$$\tilde{f}(k_n) = \int_{-L/2}^{L/2} dx f(x)e^{-ik_nx} \quad . \tag{A.11}$$

Assuming that $f(k_n)$ is a smooth function of its argument, then in Eq. (A.10) the sum can be replaced by an integral, by noting that

$$\sum_n \frac{2\pi}{L} \rightarrow \sum_n \delta k \rightarrow \int dk \quad . \tag{A.12}$$

If we let $L \rightarrow \infty$ the fourier series then turns into the fourier transform

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\tilde{f}(k)e^{ikx} \tag{A.13}$$

$$\tilde{f}(k) = \int_{-\infty}^{\infty} dx f(x)e^{-ikx} \quad . \tag{A.14}$$

The extension to three dimensions is then obvious – replace the scalars by vectors.

To return to the reciprocal lattice; Suppose we decided to take the Fourier transform of a function periodic on a lattice, i.e. $f(\mathbf{r}+\mathbf{R}) = f(\mathbf{R})$ where $\mathbf{R}$ is any vector on the Bravais lattice with primitive vectors a_i. Eq. (A.14) would then become

$$\begin{aligned}\tilde{f}(\mathbf{k}) &= \int_{\text{crystal}} d\mathbf{r} f(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}} \\ &= \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot\mathbf{R}} \int_{\text{unit cell}} d\mathbf{r}' e^{-i\mathbf{k}\cdot\mathbf{r}'} \quad .\end{aligned} \tag{A.15}$$

The sum in the last equation is over *all* lattice sites $\mathbf{R}$; only for special values of $\mathbf{k}$ where the phases in the exponential will be multiples of 2π will these sum to a value that scales proportionately to N, the number of unit cells in the system. The special values are just that $\mathbf{k} = \mathbf{G}$, with $\mathbf{G}$ a vector in the reciprocal lattice, defined above; the fourier transform reverts to a fourier series.

A note on notation. We need to be careful about counting states, on occasion, so although the distance between k-points is very small, it is worth remembering that there the density of k-points scales as 1/volume. One shuffles between sums and integrals as

$$\sum_k = \frac{(2\pi)^3}{V}\int d^3k \quad . \tag{A.16}$$

Note also that the conventional normalisations are such that the coefficients of a fourier series have the same dimension as the function; the fourier transform has the further dimensional factor of 1/Volume.

A.1.3 Delta functions

With the discrete fourier series, and the continuous fourier transform come two types of δ-functions: Kronecker, and Dirac. In both cases, they arise from *orthogonality* of the basis states.

The normalised Fourier mode basis set on a line of length a is

$$\phi_n(x) = \sqrt{\frac{1}{a}} e^{i2\pi nx/a} \quad , \tag{A.17}$$

and satisfies

$$\langle \phi_m | \phi_n \rangle = \int_{-a/2}^{a/2} dx \phi_m^*(x) \phi_n(x) = \delta_{nm} \tag{A.18}$$

This would be true for any orthonormal basis set. Invariably the δ function will appear inside a summation – for example this is how one gets the coefficients of the Fourier series, e.g.

Given the expansion

$$f(x) = \sum_n b_n \phi_n(x) \quad , \tag{A.19}$$

we calculate the coefficients b_n by multiplying both sides by $\phi_m^*(x)$ and integrating over x.

$$\begin{aligned} \int dx \phi_m^*(x) f(x) &= \int dx \phi_m^*(x) \sum_n b_n \phi_n(x) \\ &= \sum_n b_n \delta_{nm} \\ &= b_m \end{aligned} \tag{A.20}$$

which, for the case of a plane wave basis, is Eq. (A.2).

The generalisation of this to the fourier transform is that

$$\langle k | k' \rangle = \int dx e^{-ikx} e^{-ik'x} = 2\pi \delta(k - k') \quad . \tag{A.21}$$

Similarly, the Dirac δ function will always appear inside an integral, and its properties are

$$\int dx g(x) \delta(x - x_o) = g(x_o) \tag{A.22}$$

where $g(x)$ is some sufficiently smooth function.

A.1.4 Wave equations, dispersion relations, and response functions

The Fourier transform of df/dx is $ik\tilde{f}(k)$, and in general

$$\int dx \frac{d^n f(x)}{dx^n} e^{-ikx} = (ik)^n \tilde{f}(k) \tag{A.23}$$

This important result means that linear partial differential equations can be turned into *algebraic* equations in fourier space.

Consider the one-dimensional diffusion equation

$$-D\frac{\partial^2 f(x,t)}{\partial x^2}+\frac{\partial f(x,t)}{\partial t}=0 \quad . \tag{A.24}$$

Usually, one is taught to solve these things by substituting in a wave-like form $e^{i(kx+\omega t)}$, and finding the condition that allows this to be a solution; what this actually amounts to is just taking the Fourier transform of the whole equation in both space and time variables, i.e.

$$f(k,\omega)=\int dx\int dt e^{-ikx}e^{-i\omega t}f(x,t) \tag{A.25}$$

Applied to the diffusion equation above, this gives

$$(Dk^2+i\omega)f(k,\omega)=0 \quad , \tag{A.26}$$

so that there is no solution (f vanishes) unless there is a specific relationship between ω and k, called a *dispersion relation.*

This means that the general solution of the equation must be of the form[1]

$$f(x,t)=\frac{1}{2\pi}\int dk\tilde{f}(k)e^{(ikx-Dk^2t)} \tag{A.27}$$

This could of course have been written

$$f(x,t)=\frac{1}{4\pi^2}\int dk\int d\omega f(k,\omega)e^{ikx}e^{i\omega t}\delta(\omega-iDk^2) \ . \tag{A.28}$$

Often the problems are presented as an *initial value* problem, where you know the solution (and a time derivative, if necessary) at one time (say $t=0$), and are asked to *propagate* it forward in time. The general method is then as follows.

Take the initial values f(x, t=0) (and derivatives if necessary) and use the inverse transform to Eq. (A.27) (or the equivalent) to determine $\tilde{f}(k)$. Hence

$$f(k)=\int dy f(y,0)e^{iky} \quad , \tag{A.29}$$

and after substitution in Eq. (A.27) we get

$$f(x,t)=\frac{1}{2\pi}\int dk\int dy\ e^{ik(x-y)}e^{-Dk^2t}f(y,0) \quad , \tag{A.30}$$

which can be written

$$f(x,t)=\int dy\ G(x-y,t)f(y,0) \quad , \tag{A.31}$$

---

[1] In general there may be more that one solution of the dispersion relation, so there will be more modes

with the *Green's function* or *propagator* defined by

$$\begin{aligned} G(x,t) &= \frac{1}{2\pi}\int dk\; e^{ikx}e^{-Dk^2t} && \text{(A.32)} \\ &= \frac{1}{2(\pi D t)^{1/2}}e^{-\frac{x^2}{4Dt}} && \text{(A.33)} \end{aligned}$$

Another common type of problem is not free propagation, but a driven system, epitomised by the diffusion equation with a *source* term, viz.

$$D\frac{\partial^2 f(x,t)}{\partial x^2} - \frac{\partial f(x,t)}{\partial t} = c(x,t) \quad . \tag{A.34}$$

The solution to this is again straightforward using transforms:

$$f(k,\omega) = \frac{c(k,\omega)}{(Dk^2 + i\omega)} \tag{A.35}$$

The problem has been essentially solved by this step; of course if one need to have the solution as a function of space and time, there are some messy fourier transforms to do.

Here they are: in real space

$$f(x,t) = \frac{1}{4\pi^2}\int dk \int d\omega \int dx' \int dt' \frac{e^{i\omega(t-t')}e^{ik(x-x')}}{Dk^2 + i\omega}c(x',t') \quad . \tag{A.36}$$

After the ω integral is done by contour integration this can be further rewritten in terms of the Green's function Eq. (A.32)

$$f(x,t) = \int dx'\; dt'\; G(x-x', t-t')c(x',t') \quad . \tag{A.37}$$

Of course

$$G(x,t) = \int dk \int d\omega\; G(k,\omega)e^{ikx}e^{i\omega t} \quad , \tag{A.38}$$

where for this problem

$$G(k,\omega) = \frac{1}{Dk^2 + i\omega} \quad . \tag{A.39}$$

To recap:

- The Green's function, gives both the *propagation* of waves forward (or backward, for that matter - be careful!) in time and space - for this reason it is often called the propagator.
- The Green's function is also the response function to an external source.
- For a linear PDE with constant coefficients, it is trivial to write down the response function in fourier (k,ω) space, and more complicated in real space.
- The response function has a *pole* at the (possibly complex) frequency that satisfies the dispersion relation.

A.2 Variational methods

The calculus of variations is used to find the stationary value of a quantity with regard to small changes (variations) of a function upon which it depends. The variational method has great importance in physics, because most of the differential equations of physics can be framed as *variational principles*.

A.2.1 Functionals

A quick reminder: a *functional* is a quantity that depends on the value of a whole *function*, say $y(x)$. The following are functionals of y

$$\tilde{y}(k) = \int_{-\infty}^{\infty} y(x) e^{ikx}\, dx\ , \qquad \max(y(x))$$

$$I[y(x)] = \int_a^b f(y(x), y'(x), x)\, dx \tag{A.40}$$

(here F is a known function of y, $y' = dy/dx$, and x) and the following are not (i.e. just functions)

$$\sin(y(x))\ , \qquad \int_a^b \exp(iky(x))\, dk$$

A function such as f depends *locally* on the values of its arguments; a *functional* such as I depends *globally* on the whole shape of the function. We often use square brackets to denote the functional dependence.

A.2.2 Variational method

Requiring I of Eq. (A.40) to be *stationary* under small variations of $y(x)$

$$\frac{\delta I}{\delta y} = 0 \tag{A.41}$$

generates *Euler's equation*:

$$\left.\frac{\partial f}{\partial y}\right|_{y',x} - \frac{d}{dx}\left.\frac{\partial f}{\partial y'}\right|_{y,x} = 0 \quad . \tag{A.42}$$

Note the difference between the partial derivatives (where the other variables are held constant) and the total derivative

$$\frac{d}{dx} = \frac{\partial}{\partial x} + \frac{dy}{dx}\frac{\partial}{\partial y} + \frac{dy'}{dx}\frac{\partial}{\partial y'}\ . \tag{A.43}$$

Proof.

Varying the choice of function $y(x)$ causes I in Eq. (A.40) to take different values. Under an arbitrary *small* change $y(x) \to y(x) + \eta(x)$, where $\eta(a) = \eta(b) = 0$, $I[y(x)] \to I + \delta I$ where

$$\begin{aligned} \delta I[y(x), \eta(x)] &= \int_a^b \left[\frac{\partial f}{\partial y} \eta(x) + \frac{\partial f}{\partial y'} \eta'(x) \right] \\ &= \int_a^b \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} \right] \eta(x)\, dx \ . \end{aligned} \tag{A.44}$$

(The second term arises from an integration by parts.) The condition that the path $y(x)$ must obey, for I to be stationary, is then expressed as

$$\delta I[y(x), \eta(x)] = 0 \qquad \forall\, \eta(x) \tag{A.45}$$

(often written as $\delta I = 0$) which, since $\eta(x)$ is arbitrary, implies *Euler's Equation* Eq. (A.42).

A.2.3 Variation under constraints

We sometimes wish to find the function y that makes stationary the functional $I[y(x)]$ under a *constraint*

$$J[y(x)] = \int_a^b g(y, y', x) = J_o\, dx \tag{A.46}$$

with J_o a constant. This is equivalent to an unconstrained variation of the functional

$$K = I - \lambda J \quad , \tag{A.47}$$

where λ is a parameter (called an *undetermined multiplier* or a *Lagrange multiplier*). λ is to be fixed by considering boundary conditions, and the value of J_o.

A.2.4 Complex functions

We shall also deal with situations where the function y is a complex function $\phi(x) = u(x) + iv(x)$ (u, v real). In this case the variational minimisation has to be done with regard to both the real and imaginary parts of the function ϕ independently, viz.

$$\frac{\delta I[u, v]}{\delta u} = \frac{\delta I[u, v]}{\delta v} = 0 \quad , \tag{A.48}$$

However, we could equally well take variations with regard to the two independent quantities $\phi = u + iv$ and $\phi^* = u - iv$, obtaining

$$\frac{\delta I[\phi, \phi^*]}{\delta \phi} = 0 \tag{A.49}$$

$$\frac{\delta I[\phi, \phi^*]}{\delta \phi^*} = 0 \tag{A.50}$$

If I is a *real* quantity, the two Euler's equations arising from Eq. (A.49) and Eq. (A.50) will be complex conjugates of each other.

A.2.5 Quantum mechanics

In quantum mechanics, we consider the functional that is the expectation value of the energy

$$E[\phi, \phi^*] = \langle\phi|\, \hat{H}\, |\phi\rangle = \int dx\, \phi^* \hat{H} \phi \quad , \tag{A.51}$$

where the wavefunctions are assumed normalised

$$J = \langle\phi|\, |\phi\rangle = \int dx\, \phi^* \phi = 1 \quad . \tag{A.52}$$

and $\hat{H}$ is the *Hamiltonian* operator.

By making stationary the quantity $E - \lambda J$ using the complex variational Eq. (A.50) we obtain the Schrödinger equation

$$\hat{H}\phi(x) = \lambda\phi(x) \quad , \tag{A.53}$$

where now the Lagrange multiplier appears as an eigenvalue. The stationary values of E are then given by the eigenvalue(s) λ.

A.3 Elementary theory of analytic functions

For some students this may be the least familiar of the explicitly mathematical background we need for the course. Complex analysis is a large and well-developed subject, but we shall need only one elementary thing in this course - the calculus of residues. For completeness, I have included the background necessary to get there.

The results we need are to do with contour integrals in the complex $z = x + iy$ plane of analytic functions (analytic meaning differentiable). The importance for this course arises because the physical necessity of *causality* forces analytic behaviour of response functions in the complex *frequency* plane. The three or so things needed for the course are the residue theorem, Cauchy's integral representation, and the treatment of simple poles on the real axis. They are briefly summarised below, and if you are happy with the summary, you need read no further.

The residue theorem

An integral around a closed contour[2] C of a function $f(z)$ that contains only isolated singularities is the sum of the residues of that function at the poles, multiplied by $2\pi i$.

$$\int_C dz\, f(z) = 2\pi i \sum_i Res f(z_i) \tag{A.54}$$

---

[2] The convention is that the direction is counterclockwise, unless otherwise stated.

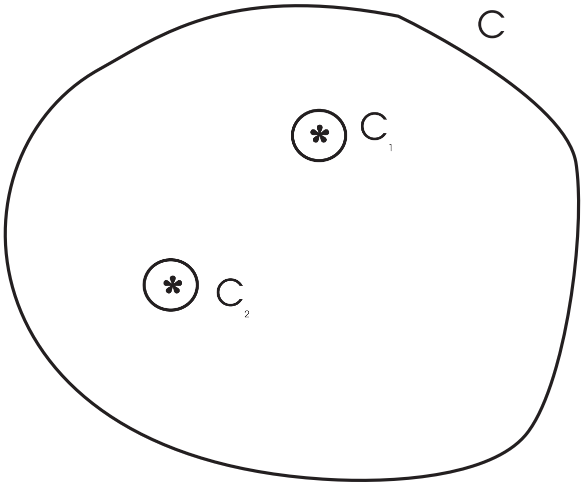


Figure A.1: A contour integral along the contour C of a function with only isolated singularities can be deformed to surround just the singularities – here at C_1 and C_2

Here the residue of the function is the coefficient d_{-1} of $1/(z - z_i)$ in the (Laurent) expansion of the function about the pole

$$f(z) = \ldots + d_2(z - z_i)^2 + d_1(z - z_i) + d_0 + \frac{d_{-1}}{z - z_i} + \frac{d_{-2}}{(z - z_i)^2} + \ldots \quad \text{(A.55)}$$

Contours can be moved around at will, as long as they stay within regions where the function is analytic and cross no singularities.

Cauchy's integral representation

A function that is analytic can be expanded as a power series (Taylor series) about a point within its radius of convergence. However, a generalisation of the residue theorem allows one to reconstruct the value of an analytic function at any point inside a closed curve, in terms of the value at the boundary.

If $f(z)$ is analytic within a closed contour C then

$$\begin{aligned} \frac{1}{2\pi i}\int_C \frac{f(z)}{z - z_0}dz &= f(z_0) \quad \text{if z}_0 \text{ is interior to } C \\ &= 0 \quad \text{if z}_0 \text{ is exterior to } C \end{aligned} \quad \text{(A.56)}$$

These results are extremely powerful, and reflect the fact that to have a function be differentiable in the complex plane produces very nonlocal constraints on its behaviour.

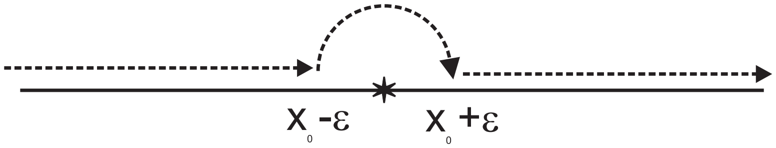


Figure A.2: Contour to evaluate the integral in Eq. (A.57)

Integrals of simple poles along the real axis

We often find ourselves needing to do an integral of the form

$$g(x_o) = \int_{-\infty}^{\infty} dx\, \frac{f(x)}{x - x_0} \tag{A.57}$$

where x, x_o are on the real axis. This needs some care, because the integral apparently runs straight through the pole. The physics will in fact always determine which way around the pole we go, so that the contour either go above, or below, the pole in the complex plane. If we take it to be as shown in Fig. A.2, we split the integral into three pieces: along the real axis up to a small distance ϵ either side of the pole, and a small semicircle round the pole.

The parts along the real line give rise to the *principal value* of the integral

$$\lim_{\epsilon \to 0} \int_{-\infty}^{x_0 - \epsilon} + \int_{x_0 + \epsilon}^{\infty} dx\, \frac{f(x)}{x - x_0} \tag{A.58}$$

If $f(x)$ is smooth near x_0, this will converge to a finite value because the positive and negative divergences of $1/(x - x_0)$ cancel. The remaining part is the integral round the semicircle, that we can get by changing variables to $z = \epsilon\, e^{i\theta}$:

$$\lim_{\epsilon \to 0} \int_{\pi}^{0} f(x_0) \frac{d(\epsilon\, e^{i\theta})}{\epsilon\, e^{i\theta}} = -i\pi f(x_0) \tag{A.59}$$

Notice how an *imaginary* part has appeared in the result, automatically due to our treatment of the pole. If we had integrated around the pole in the opposite direction, the sign of Eq. (A.59) would have been opposite.

We often express this result as shorthand

$$\frac{1}{x - x_o \pm i\eta} = \mathrm{Pr}\, \frac{1}{x - x_0} \mp i\pi\delta(x - x_0) \tag{A.60}$$

where η is an infinitesimal positive number (to remind us to keep the pole below, or above, the contour line in the complex plane as necessary), Pr stands for principal value (as in Eq. (A.58)) and $\delta(x)$ is the conventional Dirac δ-function (all of these symbols make sense only within integrals).

An outline of the theory is given in the next sections.

A.3.1 Functions of a complex variable

We shall be working in the two dimensional complex (x, y) plane, and define the complex variable

$$z = x + iy \quad .$$

We can define a function of a complex variable $f(z)$ in terms of its real and imaginary parts

$$f(z) = u(x, y) + iv(x, y) \quad . \tag{A.61}$$

In order to make use of the function $f(z)$ we need to be able to differentiate it with respect to z, which is a two dimensional coordinate; i.e. when we define

$$\frac{df}{dz} = lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \tag{A.62}$$

we must get the same answer whatever direction Δz vanishes. This produces the *Cauchy-Riemann conditions*

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} \tag{A.63}$$

$$\frac{\partial u(x, y)}{\partial y} = -\frac{\partial v(x, y)}{\partial x} \tag{A.64}$$

Proof. From Eq. (A.62) we have

$$\frac{df}{dz} = \lim_{\Delta x \to 0 \Delta y \to 0} \frac{u(x + \Delta x, y + \Delta y) + iv(x + \Delta x, y + \Delta y) - u(x, y) - iv(x, y)}{\Delta x + i\Delta y} \quad . \tag{A.65}$$

We now impose the condition that Eq. (A.65) yield the same answer independent of the order in which the limits $\Delta x, \Delta y \to 0$ are taken. If we first set $\Delta y = 0$ and take the limit $\Delta x \to 0$, we have

$$\frac{df}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial y} \quad , \tag{A.66}$$

but taking the limits in the opposite order, we have

$$\frac{df}{dz} = \frac{\partial v}{\partial x} - i\frac{\partial u}{\partial y} \quad . \tag{A.67}$$

Equating the real and imaginary parts of the two equations produces the Cauchy-Riemann conditions Eq. (A.63).

Harmonic functions. As an aside, note that if Eq. (A.63) are differentiated first with respect to x, and then with respect to y, we get

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{A.68}$$

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \tag{A.69}$$

so both the real and imaginary parts of a differentiable function satisfy Laplace's equation. This is a useful trick to solve potential problems in two-dimensions.

A.3.2 Analytic functions

A function $f(z)$ is said to be *analytic* at a given point if the function is single-valued and differentiable.

The region in the complex plane over which a function is analytic is called the *domain of analyticity*; a function that is everywhere analytic is called an *entire* function.

If a function is not analytic at some point, the point is called a *singular* point. For example $f(z) = 1/z$ is analytic everywhere except at $z = 0$ which is therefore an *isolated singular point*.

Some examples of analytic functions

(Most of these can be checked by inspection).

- A constant is an entire function.
- An integer power of z, e.g. z^n is an entire function.
- By obvious extenmsion, a polynomial *of finite order* $p_n(z) = \sum_{k=0}^{n} a_k z^k$ is also an entire function.
- A power series (a polynomial of *infinite* order)

  $$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$$

  is analytic within its radius of convergence; (i.e. the series converges for $|z - z_0| < R$).
- The exponential function e^z is an entire function — the radius of convergence of the power series

  $$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

  is infinite.
- The logarithm $\ln(z)$ is by definition the function that satisfies

  $$\exp(\ln(z)) = z$$

  If we write $z = re^{i\theta}$, we can see that in general the logarithm is multiply valued

  $$\ln(z) = \ln(r) + i(\theta + 2\pi n) \quad n = 0, \pm 1, \pm 2, ... \tag{A.70}$$

If we choose $n = 0$ so as to define a single valued function (sometimes called the *principal logarithm*, and restrict the range of θ, viz.

$$\ln(z) = \ln(r) + i\theta \quad -\pi < \theta < \pi \tag{A.71}$$

the logarithm is analytic everywhere *except on the negative real axis* where it has a discontinuity.

A.3.3 Cauchy's integral theorem

This is one of the most famous theorems of mathematics, which plays a central role in the theory of analytic functions.

If C is a closed curve in the complex plane, and a function $f(z)$ is analytic on C and everywhere within C, then the contour integral

$$\int_C dz\, f(z) = 0 \tag{A.72}$$

The importance of this theorem is that one may continuously deform a contour in an integral without changing the value of the result, provided one keeps the contour within the domain of analyticity of $f(z)$.

Sketch of proof. One way to prove the theorem is to divide up the closed area into infinitesimal squares, and replace the contour integral by a summation of contour integrals around each square (see Fig. A.3); the integrals along common boundaries of two squares cancel.

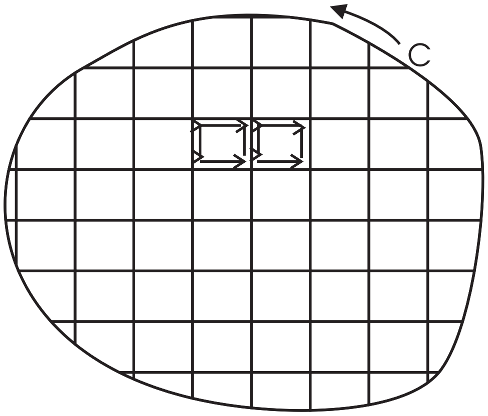


Figure A.3: Tiling of the contour for Cauchy's theorem

First, note that along a contour connecting two points z_a and z_b

$$\int_a^b z^n\, dz = \left[\frac{z^{n+1}}{n+1}\right]_a^b \tag{A.73}$$

and in particular for a closed loop $z_a = z_b$,

$$\int_a^a z^n\, dz = 0 \tag{A.74}$$

Now consider each square separately. Because $f(z)$ is analytic, one can expand the function in a power series about the centre z_o of the square, viz

$$f(z) = f(z_0) + (z - z_0)\left.\frac{df}{dz}\right|_{z_0} + ... \tag{A.75}$$

where the higher order terms will become irrelevant as the box size shrinks to zero.[3] Using Eq. (A.74), the contour integral of Eq. (A.75) around each infinitesimal square can be shown to vanish, and Eq. (A.72) follows.

A.3.4 Singular points and the residue theorem

The function $f(z) = 1/z$ is an example of a function with an isolated singular point at the origin. At an isolated singular point of a general function, one should expand the function not as a Taylor series, but as a Laurent expansion in powers of $1/(z - z_0)$, viz.

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \frac{b_1}{z - z_0} + \frac{b_2}{(z - z_0)^2} + \ldots \tag{A.76}$$

If b_1 is non-zero, but $b_2 = b_3 = \ldots = 0$, the function is said to have a *simple pole* at z_0. The coefficient b_1 is also called the *residue* of the function $f(z)$ at the pole: $b_1 = Resf(z_0)$.

Residue theorem. If a contour C contains only isolated poles at points z_i of a function $f(z)$

$$\int_C f(z)\, dz = 2\pi i \sum_i Resf(z_i) \quad . \tag{A.77}$$

Sketch of proof. We have already shown that a contour can be arbitrarily deformed through an analytic region of a function, so we know that the answer to Eq. (A.77) must just come from the singular points (see Fig. A.1)

Let us consider one singularity, conveniently placed at the origin, and we assume that

$$f(z) = \sum_{n=-\infty}^{\infty} d_n z^n$$

We deform the contour around it to C', a small circle of radius ϵ; then along the contour $z = \epsilon\, e^{i\theta}$

$$\begin{aligned} \int_{C'} f(z) dz &= \sum_{n=-\infty}^{\infty} d_n \int_0^{2\pi} \epsilon^n e^{in\theta}\, d(\epsilon e^{i\theta}) \\ &= \sum_{n=-\infty}^{\infty} d_n \epsilon^{n+1} \int_0^{2\pi} e^{i(n+1)\theta}\, d\theta \end{aligned} \tag{A.78}$$

---

[3] This is where the work goes in for the real proof

Notice that for all the terms in the sum *except* $n = -1$, the integral over θ is of a periodic function, and yields *zero*. The only term that survives is $n = -1$, hence

$$\int_{C'} f(z)dz = 2\pi i d_{-1} \tag{A.79}$$

which is conveniently *independent* of ϵ. The residue theorem Eq. (A.77) follows.

Cauchy's integral representation is a straightforward consequence of the residue theorem.

If $f(z)$ is analytic within a closed contour C then

$$\begin{aligned} \frac{1}{2\pi i}\int_C \frac{f(z)}{z-z_0}dz &= f(z_0) \quad \text{if } \mathrm{z}_0 \text{ is interior to } C \\ &= 0 \quad \text{if } \mathrm{z}_0 \text{ is exterior to } C \end{aligned} \tag{A.80}$$

Appendix B

Second quantisation

This appendix introduces some formalism for the treatment of many particle physics. It is not for examination, and except for the discussion of Heisenberg and Schrödinger representations (covered already in QM courses) it is not needed in the course.

The point of this formalism is to make descriptions of the physics, and calculations, simpler. We shall not be performing any lengthy many-body calculations, but the concepts behind the many-body formulation – particularly that of "second quantisation" – are important. The language and notation of second quantisation is all-pervasive in the research literature; the aim of this section is to help you gain some familiarity and demystify this topic.

B.1 Heisenberg and Schrödinger representations

The most familiar representation of quantum mechanics is in the wave equation, where the wave function $|\psi(t)>$ evolves in time under the presecription given by the Schrödinger equation[1]

$$i\frac{\partial|\Psi_S(t)>}{\partial t} = H|\Psi_s(t)> \quad . \tag{B.1}$$

This has a formal solution

$$|\Psi_S(t)>= e^{-iH(t-t_o)}|\Psi_S(t_o)> \quad . \tag{B.2}$$

A physical observable is represented by a time-independent operator $\hat{A}_S$, and "measurements" of this observable involve calculation of matrix elements:

$$<\hat{A}_S>=<\Psi_S|\hat{A}_S|\Psi_S> \quad . \tag{B.3}$$

---

[1] For convenience, I will set $\hbar = 1$, which means that you should think of ω and $1/t$ as energies

This is not the most convenient form to describe many-particle systems because keeping track of time-dependent wavefunctions in many variables is not straightforward. There is an alternative formulation of quantum mechanics due to Heisenberg, where the state vectors $\Psi_H >$ are time-independent and all the time-dependence is ascribed to the operators $\hat{A}_H(t)$.

Formally, this is achieved with the unitary transformation

$$\begin{aligned} |\Psi_H > &= e^{iH(t-t_o)}|\Psi_S(t)> = |\Psi_S(t_o)> \ , && \text{(B.4)} \\ \hat{A}_H(t) &= e^{iH(t-t_o)}\hat{A}_S e^{-iH(t-t_o)} \ , && \text{(B.5)} \end{aligned}$$

which evidently leaves all matrix elements – i.e. physical observables – invariant. Notice that the Hamiltonian itself is unchanged by the transformation. Instead of the Schrodinger equation, we can now computer the time-dependence of the operators by differentiating Eq. (B.5) to yield

$$i\frac{\partial \hat{A}}{\partial t} = \hat{A}H - H\hat{A} = [\hat{A}, H] \ . \tag{B.6}$$

The square bracket denotes the commutator, and Eq. (B.6) is the Heisenberg equation of motion.

B.2 Second quantisation

The next trick we need is to find a concise way of writing down a many-particle wavefunction. The complexity arises because we should use basis states which are properly (anti)-symmetrised for bosons (fermions). It was already clear from our attempts to use Slater determinants in the Hartree-Fock theory that working out a theory based on combinations of single particle wavefunctions was quite cumbersome. The process of *second quantisation*, based on the *occupation number representation* simplifies the description. Here is a brief description of the scheme[2].

We are already familiar with the idea of the Fermi sea, for noninteracting particles, where we occupied the momentum states (with spin) for momenta $k < k_F$. We never wrote down the Schrodinger wavefunction for this, instead we were actually implicitly using the *occupation number representation*, where a state is described by a vector

$$|n_1, n_2, ..., n_N > \tag{B.7}$$

which means that it contains n_1 particles in state 1, n_2 in state 2, and so on up to the state n_N. In our noninteracting Fermi sea, the state labels $1, 2, ...N$ are the momenta (and spin) $k_i, \sigma \ i = (1, N)$, and we will have $n_{k_i} = 1$ if $k < k_F$, and zero otherwise. (For fermions the occupancy is either 0 or 1, but for bosons it is unrestricted).

---

[2] For a full derivation, see either the appendix to Marder, or to Doniach and Sondheimer

What we would like to have is some algebra that will start with a given state, and add and subtract particles in a fashion that correctly preserves the symmetry or antisymmetry of the wavefunction.

B.2.1 Operators for fermions

Let us define *creation* and *annahilation* operators $\hat{c}_i^\dagger$, $\hat{c}_i$ that create or annahilate particles in the state i – which means just to change the number occupation of that state by 1. They obey the following rules:

$$\hat{c}_i|n_1 n_2 ... n_i ... > \;=\; \begin{cases} 0 & \text{if } n_i = 0 \\ |n_1 n_2 ... 0 ... > & \text{if } n_i = 1 \end{cases} \tag{B.8}$$

$$\hat{c}_i^\dagger|n_1 n_2 ... n_i ... > \;=\; \begin{cases} 0 & \text{if } n_i = 1 \\ |n_1 n_2 ... 0 ... > & \text{if } n_i = 0 \end{cases} \tag{B.9}$$

Notice that the rules clearly prevent double occupancy of a state, because the creation operator applied to a singly-occupied state gives the null vector 0; similarly since the state can never be more than singly occupied, applying the annahilation operator twice will always give 0. We must then have

$$\hat{c}_i \hat{c}_i = \hat{c}_i^\dagger \hat{c}_i^\dagger = 0 \tag{B.10}$$

One can easily check that the operation $\hat{c}_i \hat{c}_i^\dagger + \hat{c}_i^\dagger \hat{c}_i$ on an arbitrary number state recovers the state itself: hence we have

$$\hat{c}_i \hat{c}_i^\dagger + \hat{c}_i^\dagger \hat{c}_i = \{\hat{c}_i, \hat{c}_i^\dagger\} = 1 \tag{B.11}$$

where the curly brackets denote the anti-commutator. The operator $\hat{c}_i^\dagger \hat{c}_i = \hat{n}_i$ is the *number* operator, and has eigenvalues 0 or 1.

Lastly we need to consider what happens when we create and/or destroy particles in different states, i, j say. The operators carry a notion of ordering: $\hat{c}_i^\dagger \hat{c}_j^\dagger$ applied to the vacuum means to create the two particle state whose Schrödinger representation is

$$\begin{vmatrix} \phi_i(r_1) & \phi_i(r_2) \\ \phi_j(r_1) & \phi_j(r_2) \end{vmatrix} \tag{B.12}$$

The two operators applied in the opposite order must then create a state isomorphic to the determinant with the two rows interchanged, which changes the sign and expresses the principle of antisymmetry. The fermionic bookkeeping then insists upon the rule $\{\hat{c}_i^\dagger, \hat{c}_j^\dagger\} = 0$. The full set of anticommutation rules are

$$\{\hat{c}_i^\dagger, \hat{c}_j^\dagger\} = \{\hat{c}_i, \hat{c}_j\} = 0 \;\;, \;\; \{\hat{c}_i, \hat{c}_j^\dagger\} = \delta_{ij} \;\; . \tag{B.13}$$

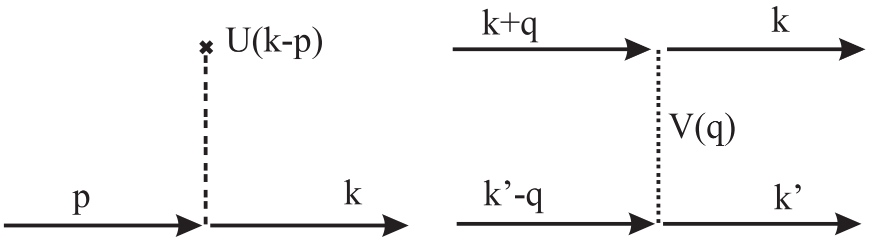


Figure B.1: Representation of potential scattering and particle-particle scattering by diagrams. Notice that momentum is conserved at each vertex

B.2.2 Hamiltonians

We now need to rewrite the Hamiltonian – and operators in general – in second quantised notation. The prescription is as follows. For a "single-particle" term - i.e. one that operates on each particle in an identical fashion we replace

$$\hat{O} = \sum_j \hat{O}(r_j) \quad , \tag{B.14}$$

by

$$\hat{O} = \sum_{kl} < k|\hat{O}|l > \hat{c}_k^\dagger \hat{c}_l \quad . \tag{B.15}$$

For example, a single particle Hamiltonian Eq. (2.12) represented in an arbitrary basis $\phi_k(r)$ becomes

$$\hat{H}_o = \sum_{kl} < k|H_o|l > \hat{c}_k^\dagger \hat{c}_l \tag{B.16}$$

where

$$< k|H_o|l >= \int dr \phi_k^*(r) \left[-\frac{\hbar^2 \nabla^2}{2m} + U(r) \right] \phi_l(r) \quad . \tag{B.17}$$

It will be often the case that we will imagine we have diagonalised the single particle H_o, and determined its eigenvalues ϵ_p and eigenstates ϕ_p; in that case, the second quantised description of H_o is also diagonal

$$\hat{H}_o = \sum_p \epsilon_p \hat{c}_p^\dagger \hat{c}_p = \sum_p \epsilon_p \hat{n}_p \tag{B.18}$$

The notion of a single-particle operator as destroying a particle in one state, and creating it in another is just that of *scattering.* For example if we represented a potential $U(r)$ using a basis of plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$, then the matrix element $< \mathbf{k}|U|\mathbf{p} >= U(\mathbf{p}-\mathbf{k})$ is just the Fourier component of the potential, that which scatters the wave from momentum $\mathbf{p}$ to $\mathbf{k}$. Such an operation is naturally represented graphically as shown in Fig. B.1.

For an operator the acts on sums of pairs of particles (usually the Coulomb interaction), the representation is a generalisation of the single particle case. We replace

$$\hat{O} = \sum_{i \neq j} \hat{O}_{ij} \tag{B.19}$$

by

$$\hat{O} = \sum_{klmn} \hat{c}_k^\dagger \hat{c}_l^\dagger \hat{c}_m \hat{c}_n < k(1)l(2)|\hat{O}_{12}|m(2)n(1) > \ , \tag{B.20}$$

where we have used the numerals $1, 2$ to refer to the particle index - note carefully the order of the operators in Eq. (B.20). For the common case of the Coulomb interaction, this matrix element is

$$< k(1)l(2)|\hat{V}_{12}|m(2)n(1) > \ = \int\int d\mathbf{r}_1 d\mathbf{r}_2 \phi_k^*(\mathbf{r}_1)\phi_l^*(\mathbf{r}_2)\frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|}\phi_m(\mathbf{r}_2)\phi_n(\mathbf{r}_1), \tag{B.21}$$

Using a basis of plane waves, the integrals in Eq. (B.21) can be evaluated, and the Coulomb interaction term is written

$$\frac{1}{2}\sum_{\mathbf{k}\mathbf{k}'\mathbf{q}} V(\mathbf{q})\hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} \ , \tag{B.22}$$

where $\hat{c}^\dagger_{\mathbf{k}\sigma}$ is the creation operator for a state of momentum $\mathbf{k}$ and spin σ. Why is there only a summation over three momenta in Eq. (B.22), whereas there are four states in Eq. (B.21)?

This interaction is often drawn as shown in Fig. B.1

B.3 Hartree-Fock revisited

I said that the reason for introducing the second quantised formulation is that it makes calculations more straightforward. Earlier we studied the Hartree-Fock theory using the cumbersome first quantised notation. Now we will look at it again using our new technology.

The Hamiltonian for the electron gas consists of the kinetic energy term plus the Coulomb interaction term of Eq. (B.22):

$$H = \sum_{\mathbf{p}\sigma}\epsilon_{\mathbf{p}}\hat{c}^\dagger_{\mathbf{p}\sigma}\hat{c}_{\mathbf{p}\sigma} + \frac{1}{2}\sum_{\mathbf{k}\mathbf{k}'\mathbf{q}\sigma\sigma'} V(\mathbf{q})\hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} \ . \tag{B.23}$$

Hartree-Fock is an example of a *mean-field theory*; we will replace products of pairs of operators by their expectation value in the fermi sea. The only expectation values that exist are

$$< \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}'\sigma'} >= n_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}')\delta_{\sigma\sigma'} \ , \tag{B.24}$$

where n_k is the occupancy of the state of momentum $\mathbf{k}$ (i.e. a number which is either 0 ($k > k_F$) or 1 ($k < k_F$))[3]. Averages such as $< \hat{c}^\dagger >$, $< \hat{c}^\dagger \hat{c}^\dagger >$ do not conserve particle number and are zero for normal systems.

We will take the four fermion operators in the interaction term of Eq. (B.23) and average them in four possible ways.

$$\begin{aligned} \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} &\approx \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} < \hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'} > \\ &- \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'} < \hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} > \\ &- \hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} < \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'} > \\ &+ \hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'} < \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} > \end{aligned} \tag{B.25}$$

In order to do the averages, we have to anticommute the operators so that they lie next to each other, which produces the sign changes in Eq. (B.25). The four terms above come from pairing: (a) the second and third operators; (b)second and fourth; (c) first and third; (d) first and fourth.

Now we use Eq. (B.24) to evaluate the averages, and the four terms in order become

$$\begin{aligned} &+ \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} n_{\mathbf{k}'}\delta(\mathbf{q}) \\ &- \hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'} n_{\mathbf{k}'}\delta(\mathbf{k}'-\mathbf{k}-\mathbf{q})\delta_{\sigma\sigma'} \\ &- \hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}+\mathbf{q}\sigma} n_{\mathbf{k}}\delta(\mathbf{k}-\mathbf{k}'+\mathbf{q})\delta_{\sigma\sigma'} \\ &+ \hat{c}^\dagger_{\mathbf{k}'\sigma'}\hat{c}_{\mathbf{k}'-\mathbf{q}\sigma'} n_{\mathbf{k}}\delta(\mathbf{q}) \end{aligned} \tag{B.26}$$

At this point, notice that if we make the permutation of variables $\mathbf{k} \to \mathbf{k}'$, $\mathbf{k}' \to \mathbf{k}$, together with $\mathbf{q} \to -\mathbf{q}$ then the first term equals the fourth, and the second term equals the third. Because all these variables are summed over in Eq. (B.23), and $V(\mathbf{q}) = V(-\mathbf{q})$ there are only two distinct terms. Putting everything back into the interaction term in the Hamiltonian, the two terms are

$$V(0)\sum_{\mathbf{k}\sigma}\hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}\sigma}\sum_{\mathbf{k}'\sigma'} n_{\mathbf{k}'} = NV(0)\sum_{\mathbf{k}\sigma}\hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}\sigma} \quad , \tag{B.27}$$

and

$$\sum_{\mathbf{q}} V(\mathbf{q})\sum_{\mathbf{k}\sigma} n_{\mathbf{k}-\mathbf{q}}\hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}\sigma} = \sum_{\mathbf{k}\sigma}\epsilon_{HF}(\mathbf{k})\hat{c}^\dagger_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}\sigma} \tag{B.28}$$

where

$$\epsilon_{HF}(\mathbf{k}) = \sum_{\mathbf{q}} V(\mathbf{q}) n_{\mathbf{k}-\mathbf{q}} = \sum_{|\mathbf{k}'|<k_F} V(\mathbf{k}-\mathbf{k}') \quad . \tag{B.29}$$

---

[3] At finite temperatures, one will have by obvious extension that $n_k = f(\epsilon_k)$ with f the fermi function.

Figure B.2: Diagrammatic representation of the direct and exchange energies corresponding to Eq. (B.27) and Eq. (B.28).

Eq. (B.27) gives the Hartree term, formally infinite – but of course cancelled exactly by the uniform positive background charge that we must introduce to keep the system neutral. Eq. (B.28) is the *exchange self-energy* that we calculated earlier in Sec. 2.2.

The algebra is still a little messy, but one of the nice advantages of this scheme is that it can all be represented graphically in *Feynman diagrams*. After a while, one learns to think in terms of these diagrams, and their ubiquity in research is one of the reasons for exposing this topic in an advanced course.

Notice that the principal effect of the averaging done in Eq. (B.24) is to fix the momenta and spin degrees of freedom. The scattering diagram of Fig. B.1 is easily visualised as a collision between two incident particles of momenta $\mathbf{k}+\mathbf{q}$ and $\mathbf{k}'-\mathbf{q}$, transferring momentum $\mathbf{q}$ between them so that the outgoing particles have momenta $\mathbf{k}$ and $\mathbf{k}'$. However, because this is a many-body system there is no sense in which the incoming and outgoing particles are the same. The creation and annahilation operators do the book-keeping, by destroying particles in the initial states, and creating particles in the final states. When we perform a mean-field theory, we wish to represent the two-particle scattering by an effective one-particle potential; that means we should focus just on one incoming and one outgoing particle, and average over the configurations of the other. On average, there is no net momentum transfer, which is why the final result is diagonal in $\mathbf{k}$. The two terms Eq. (B.27) and Eq. (B.28) come from the two topologically distinct choices for the identity of the incoming and outgoing particles; the averaging procedure can be represented by joining together the internal lines, conserving (as before) momentum and energy at each vertex, associating an internal electron line or closed loop with the factor n_k, and integrating over all the remaining variables. This process is shown diagrammatically in Fig. B.2. These diagrams are not just pictures, but they can be turned into rules for identifying and evaluating terms in many-body perturbation theory.

B.4 Particles and quasiparticles

The concept of a single electron in a many-body system is not a clear one, but we can ask much more precise questions about single-electron-like excitations. One of the quantities we should be interested in is the probability that a particle can be transported from place to place, and so we introduce the notion of the *Green's function*.

B.4.1 The Green's function

Let us imagine that we insert a particle into some state $\mathbf{p}$ at time $t = 0$, and remove a particle from a state $\mathbf{p}'$ at a later time t. The probability that a particle can propagate from one configuration to the other clearly related to the following expectation value

$$G(\mathbf{p}', t; \mathbf{p}, 0) = -i < \Psi_G | T[\hat{c}_{\mathbf{p}'}(t)\hat{c}^\dagger_{\mathbf{p}}(0)] | \psi_G > \ , \tag{B.30}$$

where Ψ_G is the many-body ground state wavefunction, and the graphical description of Eq. (B.30) is shown in Fig. B.3. T is a *time-ordering operator*, which is here to preserve causality - we shall see why in a moment. It does the following

$$\begin{aligned} T[\hat{c}(t)\hat{c}^\dagger(t')] &= \hat{c}(t)\hat{c}^\dagger(t') \quad (t > t') \ , \\ &= -\hat{c}^\dagger(t')\hat{c}(t) \quad (t' > t) \ , \end{aligned} \tag{B.31}$$

i.e. it reverses the order (changing the sign for anticommutation) depending on the order of the two times. The particle creation and annahiliation operators are in the Heisenberg representation, so

$$\hat{c}(t) = e^{iHt}\hat{c}e^{-iHt} \ . \tag{B.32}$$

where we have chosen the arbitrary zero of time to be $t = 0$, and we write $\hat{c}(0) = \hat{c}$.

Let us now work this out for a single free particle.

Using the Hamiltonian Eq. (B.18), and the Heisenberg equation of motion Eq. (B.6) the Heisenberg operator $\hat{c}(t)$ satisfies the following equation of motion

$$i\frac{\partial}{\partial t}\hat{c}_{\mathbf{p}}(t) = \epsilon_{\mathbf{p}}\hat{c}_{\mathbf{p}}(t) \ , \tag{B.33}$$

and has the solution

$$\hat{c}_{\mathbf{p}}(t) = \hat{c}_{\mathbf{p}}e^{-i\epsilon_{\mathbf{p}}t} \ . \tag{B.34}$$

What is the corresponding solution for the creation operator?

Now substitute the solution into the formula for the free particle Green's function:

$$G_o(\mathbf{p}', t; \mathbf{p}, 0) = -i < 0 | T[\hat{c}_{\mathbf{p}'}(t)\hat{c}^\dagger_{\mathbf{p}}(0)] | 0 > \ , \tag{B.35}$$

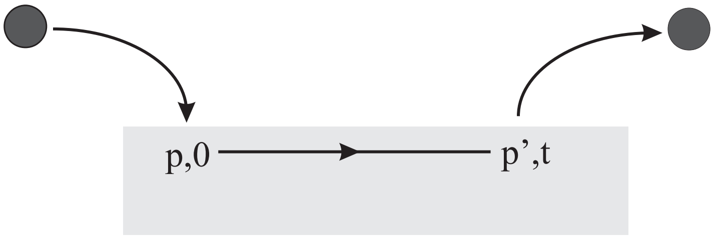


Figure B.3: Schematic picture of the Greens function $G(\mathbf{p}', t; \mathbf{p}, 0)$ for the injection of a particle in the state $\mathbf{p}$ at $t = 0$ and its removal from the state $\mathbf{p}'$ at time t. One can think of this as the propagation of a quasiparticle between the two states. The hatching represents the Fermi sea – as a reminder that the particle injected and the particle removed will not be the same in a many-body system

where now the ground state is the vacuum $|0>$. For $t > 0$, we get

$$\begin{aligned} G_o(\mathbf{p}', t; \mathbf{p}, 0) &= -ie^{-i\epsilon_{\mathbf{p}} t} < 0|\hat{c}_{\mathbf{p}'}\hat{c}^\dagger_{\mathbf{p}}|0 > \\ &= -ie^{-i\epsilon_{\mathbf{p}} t} < 0|(1 - \hat{c}^\dagger_{\mathbf{p}}\hat{c}_{\mathbf{p}})|0 > \delta_{\mathbf{p}'\mathbf{p}} \\ &= -ie^{-i\epsilon_{\mathbf{p}} t}\delta_{\mathbf{p}'\mathbf{p}} \quad , \end{aligned} \tag{B.36}$$

while for $t < 0$, the annahilation operator is to the right $G_o = 0$. Now we see the role of the time-ordering operator T - it is just to enforce causality. (This is a physical constraint, not built in to quantum mechanics which will run as well backwards as forwards in time). We also note that because of conservation of momentum, the particle only propagates within the same momentum eigenstate $\mathbf{p}$. Eq. (B.36) is precisely equivalent to the Schrodinger representation of the time-dependence of the wavefunction

$$\psi_S(t) = e^{-i\epsilon_p t}\psi_S(0) \tag{B.37}$$

We shall often be interested in the Fourier transform of G

$$G(\mathbf{p}, \omega) = \int_{-\infty}^{\infty} G(\mathbf{p}, t)e^{i\omega t}\, dt \quad , G(\mathbf{p}, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\mathbf{p}, \omega)e^{-i\omega t}\, d\omega \quad . \tag{B.38}$$

and we have therefore for the free particle propagator

$$G_o(\mathbf{p}, \omega) == \frac{1}{\omega - \epsilon_{\mathbf{p}} + i\eta} \tag{B.39}$$

where η is an infinitesimal positive constant.

The factor η sneaked into Eq. (B.39) is there for a good reason - causality again. G_o is a retarded response function, and the factor η makes clear that the pole belongs *below* the axis. There are two ways to look at this. One is in the Fourier tranform

$$\begin{aligned} G_o(\mathbf{p},\omega) &= -i\int_0^\infty dt\ e^{i(\omega-\epsilon_{\mathbf{p}})t} \\ &= -\frac{1}{\omega-\epsilon_{\mathbf{p}}}\ e^{i(\omega-\epsilon_{\mathbf{p}})t}\Big|_0^\infty \end{aligned} \tag{B.40}$$

where you will notice that we have a problem with the upper limit of the integral. If we make the change $\omega \to \omega + i\eta$, with $\eta = 0^+$, then the contribution of the upper limit vanishes as $t \to \infty$. We get back to Eq. (B.39). The other way is to look at the inverse Fourier transform, which makes the result look less like a trick and reminds one of the general principle that led to the Kramers-Krönig relations.

B.4.2 Spectral function

Returning to our discussion of the Green's function, notice that we can separate the free-particle Green's function into its real and imaginary parts

$$G_o(\mathbf{p},\omega) = \wp\frac{1}{\omega-\epsilon_{\mathbf{p}}} - i\pi\delta(\omega-\epsilon_{\mathbf{p}}) \tag{B.41}$$

The imaginary part of the Green's function is called the *spectral* function:

$$A(\mathbf{p},\omega) = -\frac{1}{\pi}\Im G(\mathbf{p},\omega)\ , \tag{B.42}$$

and for non-interacting particles

$$A_o(\mathbf{p},\omega) = \delta(\omega-\epsilon_{\mathbf{p}})\ . \tag{B.43}$$

A has a straightforward interpretation: It is the probability of finding a quasiparticle excitation of momentum $\mathbf{p}$ and energy ω; for noninteracting particles we just find excitations at the band-structure energy.

What about interacting systems? Obviously the Green's function must be different, but if the effects of interactions are somehow weak, we may expect the full Green's function to be not so different from that for free particles. One can guess what the answer ought to look like. We know that if we have a single level (our injected particle) interacting with a continuum of levels (the Fermi sea), elementary quantum mechanics tells us that the state should decay. In first quantised notation we would expect to find a wavefunction, which at long times should look like

$$\Psi_S(t) \propto e^{i\epsilon_{\mathbf{p}}t}e^{-\Gamma(\mathbf{p})t} \tag{B.44}$$

where Γ is the decay rate for the particle. Provided $\Gamma \ll \epsilon_{\mathbf{p}}$, we would say that the particle is still well-defined. Correspondingly, we might expect for the Green's function

$$G(\mathbf{p},t) = -iZ_{\mathbf{p}}e^{i\epsilon_{\mathbf{p}}t}e^{-\Gamma(\mathbf{p})t}\theta(t) \tag{B.45}$$

(here $\theta(t) = 1$ for $t > 0$ and $\theta(t) = 0$ for $t < 0$). which in frequency space is

$$G(\mathbf{p}, \omega) = \frac{Z_{\mathbf{p}}}{\omega - \epsilon_{\mathbf{p}} + i\Gamma(\mathbf{p})} \quad . \tag{B.46}$$

$Z_{\mathbf{p}}$ is an amplitude for the quasiparticle to survive to long times, and so $Z_{\mathbf{p}} \leq 1$, in general.

The corresponding spectral function is

$$A(\mathbf{p}, \omega) = \frac{1}{\pi} \frac{Z_{\mathbf{p}} \Gamma(\mathbf{p})}{(\omega - \epsilon_{\mathbf{p}})^2 + \Gamma(\mathbf{p})^2} \tag{B.47}$$

which is a Lorentzian (see Fig. 4.6) that has a total weight

$$\int_{-\infty}^{\infty} A(\mathbf{p}, \omega) d\omega = Z_{\mathbf{p}} \tag{B.48}$$

hence the factor $Z_{\mathbf{p}}$ is called the spectral weight. It turns out that one can show rigorously that the integral in Eq. (B.48) should be exactly unity; the spectral function that we have written down is valid only for long times (i.e. low frequencies) and there must be a high frequency piece that has been neglected. Eq. (B.47) defines the properties of a quasiparticle, which has a well-defined energy (that will however be different from the non-interacting energy), a lifetime $1/\Gamma$, a spectral weight Z, and satisfies Fermi statistics.