

Physics 243 - Lecture Notes

Chetan Nayak

January 27, 2004

Contents

1	Impurities in Solids	1
1.1	Impurity States	1
1.2	Density of states***	3
1.3	Localization	3
1.3.1	Anderson Model	3
1.3.2	Locator expansion***	5
1.3.3	Anderson Insulators vs. Mott Insulators	5
1.4	Physics of the Insulating State	7
1.4.1	Variable Range Hopping	7
1.4.2	AC Conductivity	9
1.4.3	Effect of Coulomb Interactions	10
1.4.4	Magnetic Properties	12
1.5	Physics of the Metallic State	16
1.5.1	Disorder-Averaged Perturbation Theory	16
1.5.2	Lifetime, Mean-Free-Path	18
1.5.3	Conductivity	20
1.5.4	Diffusion	23
1.5.5	Weak Localization	28
1.5.6	Quasiclassical Approach to Weak Localization***	33
1.5.7	Weak Magnetic Fields and Spin-orbit Interactions: the Unitary and Symplectic Ensembles	33
1.6	The Metal-Insulator Transition	33
1.6.1	Percolation	33
1.6.2	Mobility Edge, Minimum Metallic Conductivity	35

1.6.3	Scaling Theory	37
-------	--------------------------	----

CHAPTER 1

Impurities in Solids

1.1 Impurity States

In the previous parts of this book, we have discussed the low-energy excitations which result from broken symmetry, criticality, or fractionalization. In this final part of the book, we discuss the low-energy excitations which result from the presence of ‘dirt’ or ‘disorder’ in a solid. By ‘dirt’ or ‘disorder’, we mean impurities which are frozen into the solid in some random way. Consider phosphorous impurities in silicon. Presumably, the true ground state of such a mixture is one in which the phosphorus atoms form a superlattice within the silicon lattice. However, this equilibrium is never reached when the alloy is made: it is cooled down before equilibrium is reached, and the phosphorus impurities get stuck (at least on time scales which are relevant for experiments) at random positions. These random static spatial fluctuations have interesting effects on the electronic states of the system: they can engender low-lying excitations and they can dramatically change the nature of such excitations. To see the significance of this, recall that, in the absence of a broken continuous symmetry, a system will generically form a gap between the ground state and all excited states. By tuning to a critical state – such as a Fermi liquid – we can arrange for a system to have low-energy excitations. However, in the presence of disorder, it will have low-energy excitations even without any tuning. To get a sense of why this should be so, suppose that, in the absence of disorder, there were a gap

to the creation of a quasiparticle. In the presence of disorder, the potential varies from place to place, and, in the thermodynamic limit, there will be some place in the system where the potential energy offsets the energy required to create a quasiparticle. In this region, it won't cost any energy to create a quasiparticle. Note that such an excitation, though low in energy, may not have large spatial extent. By the same token, if the system were gapless in the absence of disorder, as in the case of a Fermi liquid, then disorder can extend the critical state into a stable phase and it can cause the low-lying excitations of the system to become spatially localized.

Consider the simple example of a single phosphorus impurity in silicon. The phosphorus takes the place of one of the silicon atoms. Without the phosphorus, silicon is an insulator with a band gap E_g . Phosphorus has an extra electron compared to silicon and an extra positive charge in its nucleus. Let us neglect electron-electron interactions and write Schrödinger's equation for the extra electron in the form:

$$(H_{\text{lattice}} + H_{\text{impurity}}) \psi(\mathbf{r}) = E \psi(\mathbf{r}) \quad (1.1)$$

where H_{lattice} is the Hamiltonian of a perfect lattice of silicon ions and H_{impurity} is the potential due to the extra proton at the phosphorus site. Let us write $\psi(\mathbf{r})$ in the form

$$\psi(\mathbf{r}) = \chi(\mathbf{r}) u_{\mathbf{k}_0}(\mathbf{r}) e^{i\mathbf{k}_0 \cdot \mathbf{r}} \quad (1.2)$$

where $u_{\mathbf{k}_0}(\mathbf{r}) e^{i\mathbf{k}_0 \cdot \mathbf{r}}$ is the eigenstate of H_{lattice} at the conduction band minimum. Then $\chi(\mathbf{r})$ satisfies the equation:

$$\left(-\frac{\hbar^2}{2m^*} \nabla^2 - \frac{e^2}{\epsilon r} \right) \chi(\mathbf{r}) = E_b \chi(\mathbf{r}) \quad (1.3)$$

where m^* is the effective mass in the conduction band, E_b is measured from the bottom of the conduction band, and ϵ is the dielectric constant of silicon.

Hence, so long as we can neglect electron-electron interactions, the electron will be trapped in a bound state at the impurity. If the binding energy is much less than the band gap, $\frac{e^2}{2\epsilon a_B} \ll E_g$, then our neglect of electron-electron interactions is justified because the interaction will be too weak to excite electrons across the gap. (Note that a_B is the effective Bohr radius in silicon, $a_B = \frac{\epsilon \hbar^2}{m^* e^2} \approx 20 \text{ \AA}$). Hence, in the presence of impurities, there are states within the band gap. If there is a random distribution of phosphorus impurities, we expect a distribution of bound state energies so that the gap will be filled in. In other words, there will generically be states at

the chemical potential (although there will be a small density of states when the impurity density is small), unlike in a pure system, where it is possible for the chemical potential to lie within an energy gap.

1.2 Density of states***

1.3 Localization

1.3.1 Anderson Model

What is the nature of these electronic states when there are several impurities? Presumably, there is some mixing between the states at different impurities. One can imagine that this mixing leads to the formation of states which are a superposition of many impurity states so that they extend across the system. As we will see, this naive expectation is not always correct. Consider, first, the case of a high density of electrons and impurities. One would expect the kinetic energy to increase with the density, n , as $n^{2/d}$ while the potential energy should increase as $n^{1/d}$, so that the kinetic energy should dominate for large n . When the kinetic energy dominates, we might expect some kind of one-electron band theory to be valid. This would lead us to expect the system to be metallic. How about the low-density limit? In the case of a single impurity, the electron is trapped in a hydrogenic bound state, as we saw in the previous section. What happens when there is a small, finite density of impurities? One might expect exponentially small, but finite overlaps between the hydrogenic bound states so that a metal with very small bandwidth forms. This is not the case in the low-density limit, as we will see in this section.

Consider the Anderson model:

$$H = \sum_i \epsilon_i c_i^\dagger c_i - \sum_{ij} t_{ij} c_i^\dagger c_j + \text{h.c.} \quad (1.4)$$

In this model, we ignore electron-electron interactions. Spin is an inessential complication, so we take the electrons to be spinless. ϵ_i is the energy of an electron at impurity i , and t_{ij} is the hopping matrix element from impurity i to impurity j . These are random variables which are determined by the positions of the various impurities.

One could imagine such a model arising as an effective description of the many-impurity problem. Since the impurities are located at random positions, there tunneling matrix elements between their respective bound

states will be random t_{ij} . One might also suppose that their different locations will lead to different effective environments and therefore different bound state energies ϵ_i . Anderson simplified the problem by arranging the sites i on a lattice and setting $t_{ij} = t$ for nearest neighbor sites and zero otherwise. The effect of the randomness is encapsulated in the ϵ_i 's, which are taken to be independent random variables which are equally likely to take any value $\epsilon_i \in [-W/2, W/2]$. This is a drastic simplification, but it already contains rich physics, as we will see.

For $W = 0$, there is no randomness, so the electronic states are simply Bloch waves. The system is metallic. For $W \ll t$, the Bloch waves are weakly scattered by the random potential. Now consider the opposite limit. For $t = 0$, all of the eigenstates are localized at individual sites. In other words, the eigenstates are $|i\rangle$, with eigenenergies ϵ_i . The system is insulating. Where is the transition between the $t/W = \infty$ and the $t/W = 0$ limits? Is it at finite $t/W = \infty$? The answer is yes, and, as a result, for t/W small, the electronic states are localized, and the system is insulating. (We say that a single-particle state is *localized* if it falls off as $e^{-r/\xi}$. ξ is called the *localization length*)

To see why this is true, consider perturbation theory in t/W . The perturbed eigenstates will, to lowest order, be of the form:

$$|i\rangle + \sum_j \frac{t}{\epsilon_i - \epsilon_j} |j\rangle \quad (1.5)$$

Perturbation theory will be valid so long as the second term is small. Since the ϵ_i 's are random, one can only make probabilistic statements. The typical value of $\epsilon_i - \epsilon_j$ is $W/2$. The typical smallest value for any given i is $W/2z$, where z is the coordination number of the lattice. Hence, we expect corrections to be small – and, hence, perturbation theory to be valid – if $2tz/W < 1$. On the other hand, this is not a foolproof argument because there is always some probability for $\epsilon_i - \epsilon_j$ small. Nevertheless, it can be shown (Anderson; Frohlich and Spencer) that perturbation theory converges with a probability which approaches 1 in the thermodynamic limit. Hence, there is a regime at low density (small t) where the electronic states are localized and the system is insulating, by which we mean that the DC conductivity vanishes at $T = 0$. From our discussion of the hydrogenic bound state of an electron at an impurity in a semiconductor, it is not surprising that there are localized states in a disordered system. What *is* surprising is that if the disorder strength is sufficiently strong, then *all* states will be localized, i.e. the entire band will consist of localized states.

If the disorder strength is weaker than this limit, $2tz/W \gg 1$, then we can do perturbation theory in the random potential. In perturbation theory, the states will be Bloch waves, weakly-scattered by impurities as we discuss in the next section. Such states are called *extended states*. The perturbative analysis is correct for states in the center of the band. Near the band edges, however, the perturbative analysis breaks down and states are localized: the correct measure of the electron kinetic energy, against which the disorder strength should be measured, is the energy relative to the band edge. Thus, for weak disorder, there are extended states near the band center and localized states near the band edges. When it lies in the region of extended states, it is metallic.

When the chemical potential lies in the region of localized states, the system is insulating. To see this, we write the DC conductance, g , of a system of non-interacting fermions of size L in the following form:

$$g(L) = \frac{\overline{\delta E}}{\overline{\Delta E}} \quad (1.6)$$

δE is the disorder-averaged energy change of eigenstates at E_F when the boundary conditions are changed from periodic to antiperiodic. ΔE is the mean level spacing of eigenstates. (In order to derive such a relation analytically, one must consider disorder-averaged quantities.) If the states at E_F are localized with localization length ξ , then $\delta E \sim e^{-L/\xi}$, and the conductance will vanish in the thermodynamic limit.

This form for the conductance can be motivated as follows. The conductance is the response of the current to a change in the electromagnetic potential (which can be applied entirely at the boundary). However, the potential couples to the phase of the wavefunction, so the conductance can be viewed as the sensitivity of the wavefunction to changes of the boundary conditions. A systematic derivation proceeds from the Kubo formula.

1.3.2 Locator expansion***

1.3.3 Anderson Insulators vs. Mott Insulators

The preceding analysis shows that a disorder-induced insulating state of non-interacting fermions is stable over a finite range of hopping strengths. What about the effect of electron-electron interactions? At high density, we expect the electrons to screen the impurity potentials, so that they are actually of the form $\frac{e^2}{\epsilon r} e^{-\lambda r}$. In the high-density limit, $\lambda \rightarrow \infty$, we expect the impurities to be extremely well-screened so that there won't be any impurity

bound states. Hence, electron-electron interactions should, naively, enhance metallic behavior at high density. At low density, $\lambda \rightarrow 0$, so the impurities are not effectively screened, and we expect the preceding analysis to hold. To summarize, electron-electron interactions tend to stabilize the metallic state by screening the random potential due to impurities. However, when the density is sufficiently low, this screening effect is too weak to prevent localization. As we will discuss in a later section, unscreened electron-electron Coulomb interactions do strongly influence the properties of the insulating state.

Of course, electron-electron interactions can do other things besides screen the impurities. They can lead to ordered states such as superconductivity or ferromagnetism, which would, of course, require a re-examination of the above analysis. In fact, sufficiently strong interactions can also cause insulating behavior even in an extremely clean system. At low densities where Coulomb interactions dominate over the kinetic energy, electrons will form a Wigner crystal. A tiny amount of disorder will pin this crystal, thereby making it insulating. Hence, it would be more accurate to say that electron-electron interactions enhance metallic behavior in the high-density limit but cause insulating behavior in the low-density limit.

The term ‘Anderson insulator’ is usually used for an insulator which would be metallic in the absence of impurities. The implication is that it can be more or less continuously deformed into a non-interacting fermion system which is insulating as a result of Anderson localization. A ‘Mott insulator’ is an insulator which would be metallic in the absence of electron-electron interactions. The Wigner crystal is an example of a Mott insulator. Another canonical example is given by the Hubbard model:

$$H = - \sum_{\langle i,j \rangle} t c_{i\alpha}^\dagger c_{j\alpha} + \text{h.c.} U \sum_i n_{i\uparrow} n_{i\downarrow} \quad (1.7)$$

For $U = 0$, it would be metallic at half-filling. However, for sufficiently large U there is precisely one electron. Because the energy cost for having two electrons on any site is prohibitive, the system is insulating.

Doped semiconductors and many other real materials do not fall so neatly into one class or the other, however. While they might be metallic in the absence of both disorder and electron-electron interactions, their insulating behavior cannot be entirely ascribed to one or the other. Let us reconsider, momentarily, the formation of hydrogenic bound states in a doped semiconductor. If we were to neglect localization, we would still expect a sharp crossover between highly-conducting and poorly-conducting regimes at, respectively, high and low densities. The screened Coulomb interaction,

$V(r) \sim \frac{e^2}{\epsilon r} e^{-\lambda r}$ has no bound states in the high-density limit, where λ is large. Therefore, as the dopant density is decreased, bound states will begin to appear when λ falls below a certain value $\sim a_B$. This would lead to small conductivity. (But not zero yet because the exponentially-small overlap between these bound states would still be expected to lead to metallic behavior, albeit with a very narrow band, were it not for localization.) Anderson localization of the impurity bound states will make the crossover from a metal to a poor metal into a sharp metal-insulator transition. Thus, it is more accurate to say that insulating behavior is due to the interplay between localization and interactions.

1.4 Physics of the Insulating State

At $T > 0$, any system will have finite conductivity since there will always be some probability that thermally-excited carriers can transport charge across the system. The conductivity of a finite-sized system will also always be non-vanishing. For instance, consider a non-interacting Fermi system whose single-particle states at the Fermi energy are localized with localization length ξ . Then the conductivity of the system will be $\sigma \sim L^{2-d} e^{-L/\xi}$ in a system of size L .

Thus, we define an insulator as any state in which the DC conductivity vanishes at $T = 0$ in the thermodynamic limit. In this section, we discuss the properties of a disorder-driven insulating state.

When a solid is driven into an insulating phase as a result of impurities, it is as far from a perfect crystalline lattice as possible, as far as its electronic properties are concerned. Not only is translational symmetry absent, but physical properties are dominated by localized excitations so that it is difficult to construct even an averaged description which is translationally invariant. However, some of these properties – the DC and AC conductivities, magnetic response, and specific heat – can be deduced by simple arguments, which we present in this section.

1.4.1 Variable Range Hopping

One of the most celebrated results in the study of disordered insulators is Mott's formula for the DC conductivity due to *variable range hopping*. Let us consider an insulator at low temperatures. We imagine that the electrons are all in single-particle localized states. However, at finite-temperature, an electron can be excited by, say, a phonon so that it hops to another nearby localized state. Through a sequence of such hops, it can conduct

electricity. One might imagine that electrons will only hop to the nearest neighboring states since the matrix element to hop to a more distant state will be exponentially (in distance) smaller. However, a state which is further away might be closer in energy, and therefore exponentially (in temperature) easier to reach. The competition between these competing effects leads to variable range hopping.

Suppose that an electron which is in a localized electronic state ψ hops to another one a distance R away. The number of available states per unit energy within such a distance is $\frac{4}{3}\pi R^3 N_F$, where N_F is the density of states at the Fermi energy. Hence, the typical smallest energy difference between the state ψ and a state within a distance R is the inverse of this:

$$W(R) \equiv \left(\begin{array}{c} \text{typical smallest energy difference between} \\ \text{states separated by a distance } R \end{array} \right) = \frac{3}{4\pi N_F R^3} \quad (1.8)$$

The conductivity associated with such a hopping process depends on the matrix element (squared) between the two states, which we take to be exponentially decaying $\sim e^{-2\alpha\bar{R}}$ (where $\bar{R} = 3R/4$ is the average hopping distance) and on the probability that the electron receives the activation energy needed to hop, which follows the Boltzmann distribution $\sim e^{-\beta W(R)}$. Hence, the conductivity due to such a hopping process varies as

$$\sigma_R \sim e^{-2\alpha\bar{R} - \beta W(R)} \quad (1.9)$$

The prefactor of the exponential contains various comparatively weak dependences on R ; for instance, there should be a factor of R corresponding to the fact that a longer hop constitutes a larger contribution to the current.

There are many such hopping processes available to our electron in state ψ . The electron could hop a short distance with relatively large matrix element, but the activation energy would probably be high in such a case. It could instead hop a long distance and find a state with low activation energy, but the matrix element would be very small. The optimal route is for the electron to hop a distance which minimizes the exponential in (1.9):

$$-2\alpha - \beta \frac{dW}{dR} = 0 \quad (1.10)$$

or

$$R_{\text{opt}} = \left(\frac{3}{2\pi\alpha N_F T} \right)^{1/4} \quad (1.11)$$

Hence, as the temperature is decreased, the electron makes longer and longer hops. The resulting temperature dependence of the conductivity is:

$$\sigma(T) = \sigma_0(T) e^{-(T_0/T)^{1/4}} \quad (1.12)$$

where $\sigma_0(T)$ is relatively weakly-dependent on T , e.g. power-law. In d dimensions, the preceding analysis immediately generalizes to

$$\sigma(T) = \sigma_0(T) e^{-(T_0/T)^{\frac{1}{d+1}}} \quad (1.13)$$

1.4.2 AC Conductivity

There are many situations in which the $\omega \gg k_B T$ frequency dependence of a quantity can be obtained from its DC temperature dependence by simply replacing $k_B T$ with ω . The conductivity of an insulator does not fall within this class. The AC conductivity is entirely different in form.

At frequency ω , an electron will oscillate between states separated in energy by ω . Let us think about these resonant states in terms of hydrogenic bound states. There are two possibilities. (1) R is large, so that they are essentially individual, independent bound states separated in energy by the difference between their respective potential wells. Then the dipole matrix element is $\sim R e^{-\alpha R}$; the current matrix element is $\sim \omega R e^{-\alpha R}$. Since the matrix element will be exponentially suppressed, the contribution of such resonant pairs will be small, and we turn to the second possibility. (2) The distance R is not very large, so that the energy eigenstates are actually linear combinations of the hydrogenic bound states. Then the energy difference between the eigenstates is essentially the matrix element of the Hamiltonian between the hydrogenic bound states $\sim I_0 e^{-\alpha R}$. The dipole matrix element is $\sim R$; the current matrix element is $\sim \omega R$. There is no exponential suppression because the two eigenstates are different linear combinations of the same two bound states.

Then the conductivity due to this resonant pair is:

$$\sigma(\omega) \sim (\text{dipole matrix element})^2 \cdot (\text{available phase space}) \quad (1.14)$$

The available phase space is the density of states per unit volume multiplied by the volume in which the states can lie, given that they are a distance R apart: $\sim N_F^2 R^2 \Delta R$. ΔR is the size ξ of a localized state. Hence,

$$\sigma(\omega) \sim (\omega R)^2 R^2 \sim \omega^2 R^4 \quad (1.15)$$

However, by the arguments of the preceding paragraph,

$$R \sim \ln(I_0/\omega) \quad (1.16)$$

Thus, we have the following formula for the conductivity:

$$\sigma(\omega) \sim \omega^2 \ln^4(I_0/\omega) \quad (1.17)$$

This formula also holds at finite temperatures, so long as $\omega \gg T$. However, when the frequency is decreased, the conductivity will cross over to the variable-range hopping form.

1.4.3 Effect of Coulomb Interactions

In an insulator, Coulomb interactions are not screened at long-wavelengths and low-frequencies because there aren't mobile charges. Thus, their effects are particularly strong. One effect, pointed out by Efros and Shklovskii is a suppression of the single-particle density-of-states at the Fermi energy. The basic physics is as follows. Suppose that there were two (strongly-localized) single-particle states, one just below the Fermi energy and one just above. The 'single-particle energy cost' associated with promoting an electron from the lower one to the higher one would be at least partially offset by the negative Coulomb energy associated with the resulting particle-hole pair. If the distance between the states is small, this Coulomb energy will be large and will overcompensate the single-particle energy. Hence, the states must be far apart and, therefore, the single-particle density of states (per unit volume) must be small.

More quantitatively, suppose the single-particle energy separation between an un-occupied state above the Fermi energy and an occupied one below the Fermi energy is $\Delta E_{\text{unocc-occ}}$ while the distance between the states is r . The 'single-particle energies' include the Coulomb interaction energy between the two states under consideration and all of the other occupied states in the system. Then it must be the case that

$$\Delta E_{\text{unocc-occ}} > \frac{e^2}{r} \quad (1.18)$$

or else the energy would be lowered by transferring an electron from the occupied state to the unoccupied one, contradicting our assumption. The right-hand-side is the Coulomb energy between the electron and the hole which results when such a transfer occurs.

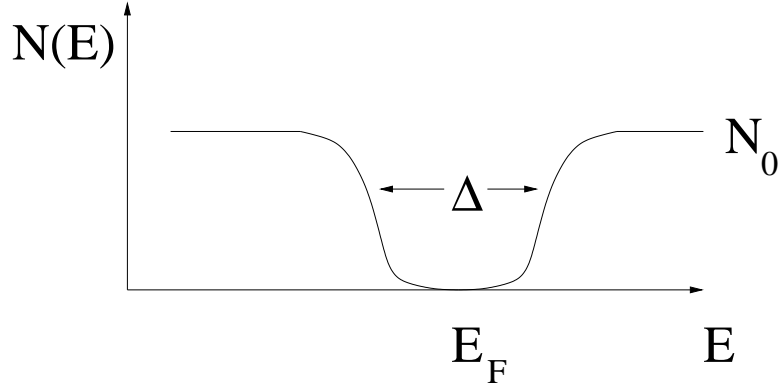


Figure 1.1: The single-particle density-of-states is suppressed at the Fermi energy. For $|E_{E_F}| > \Delta$, the single-particle density-of-states returns to N_0 .

Hence, the number of states per unit volume which are within energy ΔE of the Fermi surface is

$$n(\Delta E) \sim \frac{1}{r^3} < (\Delta E)^3 \quad (1.19)$$

from which we see that the density of states is

$$N(\Delta E) \sim (\Delta E)^2 \quad (1.20)$$

Thus, the single-particle density of states vanishes at the Fermi energy, and is suppressed near it. This effect is called the *Coulomb gap*.

Suppose we define energy, length scales Δ , r_Δ by

$$\begin{aligned} \frac{e^2}{r_\Delta} &= \Delta \\ N_0 r_\Delta^3 \Delta &= 1 \end{aligned} \quad (1.21)$$

where N_0 is the density of states in the absence of long-range Coulomb interactions, to which the density of states returns when the distance from the Fermi surface exceeds $\Delta = e^3 \sqrt{N_0}$, as shown in figure 1.1.

One could measure the single-particle density of states by performing a tunneling experiment in which current tunnels into a disordered insulator from a metallic lead to which it is contacted. The differential conductance dI/dV will be proportional to the density of states. According to the above arguments, dI/dV will be suppressed at zero voltage and will increase with

voltage as V^2 . However, when we change the applied voltage on our disordered insulator, the Fermi energy moves with it. Shouldn't the differential conductance remain zero? The answer is no. As we change the voltage, dI/dV increases with voltage as V^2 , as if the chemical potential had not moved. However, if we raise the temperature of the insulator, change the applied voltage to V_0 , and cool it down again, we find that dI/dV varies as $(V - V_0)^2$, so that the Fermi energy has moved to V_0 . Essentially, the suppression of the density of states remains wherever it was when the system last equilibrated. The non-equilibrium nature of the Coulomb gap is related to the assumption above that the only Coulomb interaction energy which must be accounted for is that between the excited electron and the hole left behind, $-e^2/r$. However, when an electron is excited from an occupied state to an unoccupied one, all of the other electrons can and will rearrange themselves. This will occur in such a way as to allow a state near the Fermi surface. Thus, if an experiment were done slowly enough to allow such re-equilibration, the suppression calculated above would not occur.

The suppression of the density-of-states near the Fermi energy has consequences for variable-range hopping. Recall that we estimated that within a region of size R , the typical smallest energy separation ΔE is $\Delta E \sim 1/(N_0 R^3)$. However, $N(\Delta E) \sim (\Delta E)^2$, or $\Delta E \sim 1/R$, as expected if the energy is determined by Coulomb interactions. Following the minimization procedure of subsection 1.4.1, we find that the variable-range-hopping formula in the presence of Coulomb interactions (in arbitrary dimension) is:

$$\sigma \sim e^{-(T_0/T)^{1/2}} \quad (1.22)$$

1.4.4 Magnetic Properties

In the absence of electron-electron interactions, there will be two electrons in each localized single-particle state, and the ground state will be paramagnetic. However, it is clearly unrealistic to ignore electron-electron interactions. We expect them to prevent two electrons from occupying the same state. Thus, we must consider the spin-spin interactions between the electrons in different localized states.

An isolated impurity state, occupied by a single electron will have a Curie form for the magnetic susceptibility:

$$\chi_c(T) \sim \frac{1}{T} \quad (1.23)$$

If we have many such spins but ignore the interactions between them, we

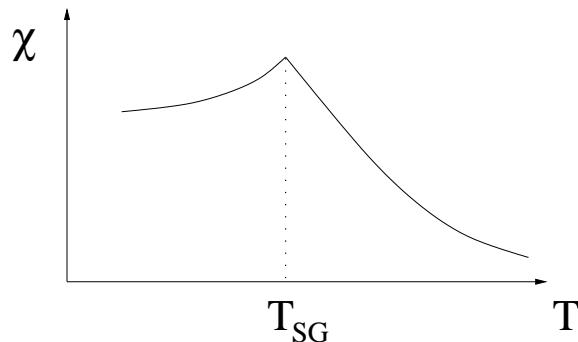


Figure 1.2: The susceptibility of a spin-glass has a kink at T_{SG} .

will have a susceptibility of the same form, proportional to the density of spins.

If, however, we include the interactions between spins, the susceptibility will have this form at high-temperatures, but once T becomes comparable to or less than the typical spin-spin interaction strength J , $\chi(T)$ will exhibit different behavior. There are several possibilities. One, which occurs in, for instance, Cu:Mn and Au:Fe alloys is *spin glass* behavior, in which the susceptibility has a cusp at a temperature T_{SG} , as shown in figure 1.2. The non-linear susceptibility actually diverges at T_{SG} . Spin glasses are rather incompletely understood, and many of their properties are still the subject of controversy. However, the basic caricature is that the spins order, but unlike in a ferromagnet, in which they all point in the same direction, the spins point in a random (but fixed) direction which varies from point to point. At the freezing temperature, T_{SG} , the system undergoes a transition into such a state in which $\langle \mathbf{S}(\mathbf{x}) \rangle = \mathbf{s}_0(\mathbf{x})$ where $\mathbf{s}_0(\mathbf{x})$ is a randomly-varying function of \mathbf{x} . There is some disagreement about whether or not such a state has infinitely-many distinct degenerate ground states (over and above the degeneracy due to broken rotational symmetry) and the related issue of whether or not there is a sharp phase transition in non-zero magnetic field. These issues are discussed in (refs.).

Spin-glass order is believed to require spin-spin interactions of both ferromagnetic and antiferromagnetic signs which are sufficiently disordered as to frustrate either of these ordering tendencies of clean systems. As a result, the spins order in a novel type of ordered state which can occur only in a dirty system. An alternate possibility which can occur in a system which has predominantly antiferromagnetic interactions (but of randomly-varying

strengths) is that the spins do not order but rather form a *random singlet phase*. In such a phase, the system essentially breaks up into a set of pairs of spins which interact with strong antiferromagnetic exchange couplings J_i . Each of these pairs has susceptibility

$$\chi_{i^{\text{th pair}}}(T) = \frac{1}{T} \frac{2}{3 + e^{\beta J_i}} \quad (1.24)$$

The susceptibility of the system must be summed over all pairs with some probability distribution, $P(J)$, for the values of J :

$$\chi(T) = \frac{1}{T} \int dJ P(J) \frac{2}{3 + e^{\beta J}} \quad (1.25)$$

At a given temperature, T , there will be some spins whose J is smaller than T . Such spins will not have formed singlets yet, so they will give a Curie-like contribution to the susceptibility, which will be the dominant contribution:

$$\chi_{\text{unpaired}}(T) \sim \frac{\rho(T)}{T} \quad (1.26)$$

where $\rho(T)$ is the number of unpaired spins at temperature T .

The function $\rho(T)$ depends on the distribution of exchange couplings, $P(J)$. Let us suppose that there is a small density n of randomly-situated localized spins, and suppose that the coupling between spins separated by a distance r is $J(r) = J_0 e^{-2r/a}$ for some constant a which is roughly the size of the localized states. With this functional form for J , we know $P(J)$ if we know $P(R)$ such that $P(R)dR$ is the probability that a spin has its nearest neighbor lying in a thin shell between R and $R + dR$. We can compute $P(R)$ in the following way. Define $p(R)$ as the probability that the nearest neighbor is a distance R or less. Then

$$P(R) = \frac{dp}{dR} \quad (1.27)$$

The probability that the nearest neighbor is between R and $R + dR$ is given by the number of impurities in a thin shell at R , $4\pi nR^2 dR$, multiplied by the probability that there is no closer neighbor, $1 - p(R)$

$$P(R) = 4\pi nR^2(1 - p(R)) \quad (1.28)$$

or, simply,

$$\frac{dp}{dR} = 4\pi nR^2(1 - p) \quad (1.29)$$

Solving this differential equation, we have

$$p(R) = 1 - e^{-4\pi n R^3} \Rightarrow P(R) = 4\pi n R^2 e^{-4\pi n R^3} \quad (1.30)$$

We substitute this expression into (1.25),

$$\chi(T) = \frac{1}{T} \int dR 4\pi n R^2 e^{-\frac{4}{3}\pi n R^3} \frac{2}{3 + e^{\beta J(R)}} \quad (1.31)$$

In the low-temperature limit, this is simply

$$\chi(T) \approx \frac{1}{T} \int dR 8\pi n R^2 e^{-\frac{4}{3}\pi n R^3} e^{-\beta J_0 \exp(-2R/a)} \quad (1.32)$$

The integrand has a saddle-point at $R \approx \frac{a}{2} \ln(J_0/T)$, dropping $\ln \ln T$ terms which are subleading at low T . Hence, the integral can be evaluated by saddle-point approximation:

$$\chi(T) \sim \frac{1}{T} e^{-\frac{1}{6}\pi n a^3 \ln^3(J_0/T)} \quad (1.33)$$

The physical interpretation of this result is simple. Unpaired spins at temperature T are those spins whose largest coupling to the other spins is less than T :

$$J(R) \sim T \Rightarrow R \sim \frac{a}{2} \ln(J_0/T) \quad (1.34)$$

The density of such spins should therefore be:

$$\rho(T) \sim e^{-\frac{1}{6}\pi n a^3 \ln^3(J_0/T)} \quad (1.35)$$

as we found above.

The approximate form which we used for $P(R)$ underestimates the number of unpaired spins, thereby underestimating the susceptibility. We took $P(R)$ to be the probability of having a neighboring spin at R , namely $4\pi n R^2$, multiplied by the probability that there were no closer neighbors, $1 - p(R)$. However, the neighboring spin at R might itself have already paired up with a third spin, in which case our spin must look for an even further neighbor with which to form a singlet pair. Thus, we should multiply by the probability that the neighbor at distance R doesn't have a closer neighbor of its own, which is roughly $1 - p(R)$. (It should really be a fraction of this because we should exclude the region which is less than a distance R from both spins since this region has already been accounted for by the first factor of $1 - p(R)$.) Hence,

$$P(R) = 4\pi n R^2 (1 - p(R))^2 \quad (1.36)$$

or

$$P(R) = \frac{4\pi n R^2}{\left(1 + \frac{4}{3}\pi n R^3\right)^2} \quad (1.37)$$

Substituting this into (1.25), we find the low-temperature form

$$\chi(T) \approx \frac{1}{T} \int dR \frac{8\pi n R^2}{\left(1 + \frac{4}{3}\pi n R^3\right)^2} e^{-\beta J_0 \exp(-2R/a)} \quad (1.38)$$

The saddle-point is again at $R \approx \frac{a}{2} \ln(J_0/T)$, up to $\ln \ln T$ terms. Hence, the saddle-point approximation for the integral is:

$$\chi(T) \sim \frac{1}{T \ln^3(J_0/T)} \quad (1.39)$$

Hence, the number of unpaired spins,

$$\rho(T) \sim \frac{1}{\ln^3(J_0/T)} \quad (1.40)$$

goes to zero very slowly; so slowly, in fact, that the susceptibility diverges at $T \rightarrow 0$.

In 1D, the asymptotic low-temperature behavior can be found exactly (refs.).

Exercise: Compute the specific heat at low T in a random singlet phase.

1.5 Physics of the Metallic State

1.5.1 Disorder-Averaged Perturbation Theory

A metallic state in a disordered system can be approached perturbatively from the clean metallic state. In order to do so, we will need to handle the perturbation caused by impurity scattering. In this section, we describe the simplest way to do so. Let us imagine breaking our system into a large number N cells, each of which is macroscopic. Then the configuration of impurities in each cell will be different. If we compute an extensive quantity, such as the free energy, then it will be essentially (up to boundary terms) the sum of the free energies of each of the cells. Since each cell will have a different impurity configuration, the sum over cells will be an average over impurity configurations. Hence, the free energy in a disordered system can be computed by computing the free energy for a fixed realization of the disorder and then averaging over all realizations of the disorder (with respect

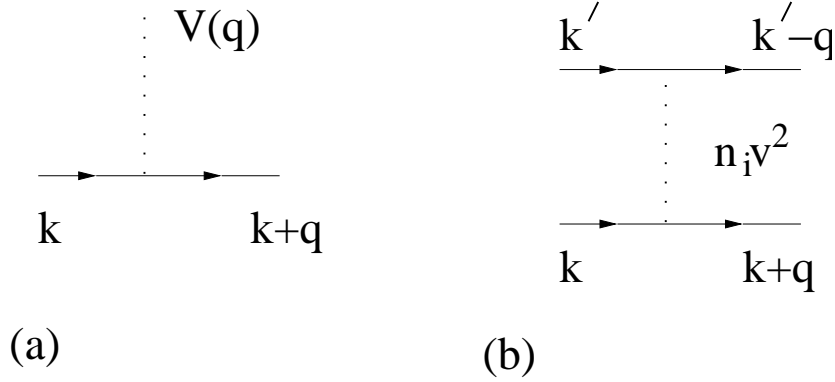


Figure 1.3: (a) The impurity vertex for a fixed $V(q)$. (b) The induced ‘interaction’ which results from averaging over the probability distribution for $V(q)$.

to some probability distribution for the disorder). Such an approximation (which we take to be exact in the thermodynamic limit) for the free energy is called the *quenched approximation* and such an average is called a *quenched average*.

Let us consider a perturbative computation of the free energy. For the sake of concreteness, suppose that the action is

$$S[\psi, \psi^\dagger] = S_{\text{clean}}[\psi, \psi^\dagger] + \int d\tau d^3x V(x) \psi^\dagger(x, \tau) \psi(x, \tau) \quad (1.41)$$

Then there is a vertex in the theory of the form depicted in figure 1.3a. To each such vertex, we assign a factor of the Fourier transform of the potential, $V(q)$ and integrate over \mathbf{q} . We can therefore organize the free energy in powers of $V(q)$:

$$F[V(q)] = \sum_n \frac{1}{n!} \int \dots \int V(q_1) \dots V(q_n) f(q_1, \dots, q_n) \quad (1.42)$$

If we think about computing the free energy perturbatively in some coupling constant in S_{clean} (which includes the simplest case, namely that S_{clean} is a free theory), then the free energy will only contain *connected diagrams*.

Now suppose that we average over $V(q)$. Let us consider, for simplicity, a distribution for $V(q)$ of the form

$$\overline{V(\mathbf{q})} = 0$$

$$\overline{V(\mathbf{q})V(\mathbf{q}')} = n_i v^2 \delta(\mathbf{q} + \mathbf{q}') \quad (1.43)$$

where n_i is the impurity concentration and v is a measure of the strength of the scattering potential due to each impurity. All higher moments are taken to be factorizable into products of $\overline{V\overline{V}}$. In other words, we will average $F[V(q)]$ over the distribution:

$$\overline{F} = \int \mathcal{D}V F[V] e^{-\int d^x V^2(x)/2v^2} \quad (1.44)$$

Integrating over V ties together different impurity vertices with impurity lines in all possible ways, according to Wick's theorem. To each such line, we attach a factor of $n_i v^2 \delta(\mathbf{q} + \mathbf{q}')$, as shown in figure 1.3b.

In essence, we have a new Feynman rule associated with impurity lines connecting impurity vertices. There are two important differences with most of the Feynman rules which we have considered thus far. (1) No energy flows along these lines, only momentum. (2) The diagrams must be connected even if all of the impurity lines are cut. In particular, if the clean theory is a non-interacting theory, then the disorder-averaged perturbation theory can have no closed electron loops. For example, figure 1.4a is allowed, but 1.4b is not.

1.5.2 Lifetime, Mean-Free-Path

The properties of a dirty metal can be derived neatly with the use of an effective field theory which we will discuss in the next chapter. Therefore, we will just summarize these properties here and describe their qualitative features.

Many of the properties of a dirty metal are qualitatively the same as those of a clean metal: the compressibility is finite and approximately temperature-independent at low-temperatures, as is the magnetic susceptibility. The specific heat is linear in T . The principal new feature in a dirty metal is that the DC conductivity does not diverge at low-temperatures; it approaches a constant value at $T = 0$.

This can be seen with a perturbative calculation. Consider the electron Green function. The lowest-order contribution to the self-energy comes from the diagram depicted in figure 1.4a. We will call the imaginary part of this diagram $(i/2\tau) \text{sgn}(\epsilon_n)$, for reasons which will be clear shortly. If the interaction between the electrons and the impurities is

$$H_{\text{imp}} = \int \frac{d^3 k}{(2\pi)^3} \frac{d^3 k'}{(2\pi)^3} \frac{d\epsilon}{2\pi} V(k' - k) \psi^\dagger(k, \epsilon) \psi(k', \epsilon) \quad (1.45)$$

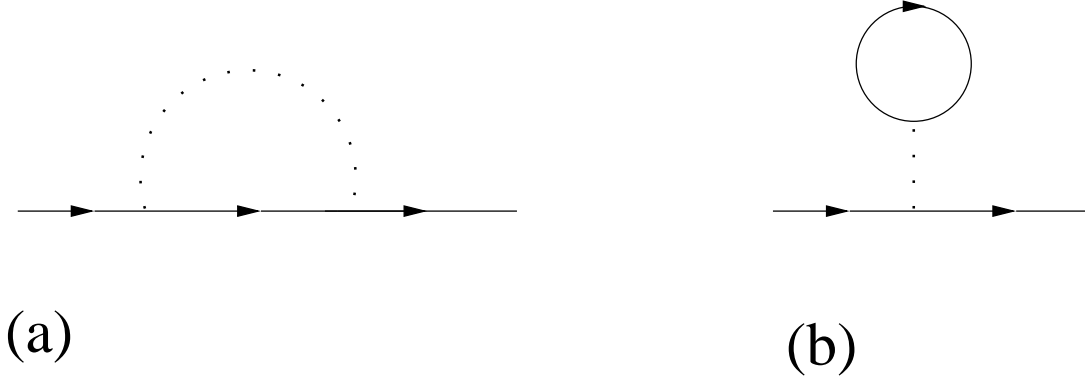


Figure 1.4: (a) An allowed diagram in disorder-averaged perturbation theory. This diagram is the lowest-order contribution to the self-energy. (b) A disallowed diagram in disorder-averaged perturbation theory.

where, for simplicity, we take

$$\overline{V(\mathbf{q})} = 0, \quad \overline{V(\mathbf{q})V(\mathbf{q}')} = n_i v^2 \delta^{(3)}(\mathbf{q} + \mathbf{q}') \quad (1.46)$$

The disorder-averaged value of this diagram is equal to

$$\begin{aligned} \frac{1}{2\tau} \text{sgn}(\epsilon_n) &\equiv \text{Im}\Sigma(\epsilon, k) \\ &= \text{Im} \left\{ \int \frac{d^3 k'}{(2\pi)^3} \frac{\overline{V(\mathbf{k} - \mathbf{k}')V(\mathbf{k}' - \mathbf{k})}}{i\epsilon_n - \xi_{\mathbf{k}}} \right\} \\ &= 2\pi N_F n_i v^2 \text{sgn}(\epsilon_n) \end{aligned} \quad (1.47)$$

The real part merely renormalizes the chemical potential.

If we add this self-energy to the inverse Green function, we have a Matsubara Green function of the form:

$$\mathcal{G}(\mathbf{k}, \epsilon_n) = \frac{1}{i\epsilon_n - \xi_{\mathbf{k}} + \frac{i}{2\tau} \text{sgn}(\epsilon_n)} \quad (1.48)$$

The corresponding retarded and advanced Green functions are:

$$G_{\text{ret,adv}}(\mathbf{k}, \epsilon) = \frac{1}{\epsilon - \xi_{\mathbf{k}} \pm \frac{i}{2\tau}} \quad (1.49)$$

In the presence of impurities, momentum eigenstates are no longer energy eigenstates. The inverse-lifetime $1/\tau$ is the energy-uncertainty of a

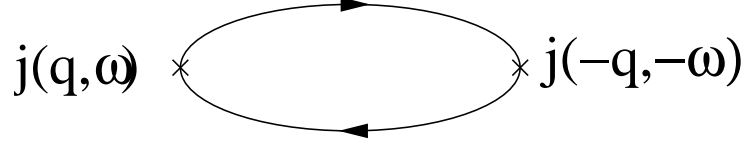


Figure 1.5: The basic conductivity bubble. The electron lines represent lines dressed by the one-loop self-energy correction computed in the previous section.

momentum eigenstate. Writing $\xi_k = v_F |k - k_F|$, we see that the mean-free-path $\ell = v_F \tau$ is the momentum uncertainty of an energy eigenstate. It is essentially the average distance between scatterings of an electron since the momentum is the inverse of the spatial rate of change of the wavefunction and the latter is scrambled at collisions.

Fourier transforming the fermion Green function at the Fermi energy into real space, we find

$$G_{\text{ret}}(\mathbf{x}, \epsilon) \sim e^{-|\mathbf{x}|/\ell} \quad (1.50)$$

Thus, the single-particle Green function decays exponentially, unlike in a clean system, where it has power-law behavior.

1.5.3 Conductivity

With this single-particle Green function, we can perturbatively compute the conductivity. Consider the basic ‘bubble’ diagram, figure 1.5, at $\mathbf{q} = 0$ but with each electron line representing the Green function (1.48) rather than the bare Green function.

$$\begin{aligned} \langle j(0, \omega_n) j(0, -\omega_n) \rangle = \\ \frac{1}{3} v_F^2 N_F \int d\xi_k \frac{1}{\beta} \sum_n \frac{1}{i\epsilon_n + i\omega_n - \xi_k + \frac{i}{2\tau_{n+m}}} \frac{1}{i\epsilon_n - \xi_k + \frac{i}{2\tau_n}} \end{aligned} \quad (1.51)$$

where we have introduced the shorthand $\tau_n = \tau \text{sgn}(\epsilon_n)$. The factor of 1/3 comes from the angular average over the Fermi surface. As usual, we convert the sum over Matsubara frequencies to an integral

$$\begin{aligned} \langle j(0, \omega_n) j(0, -\omega_n) \rangle = \\ - \frac{v_F^2}{3} N_F \int d\xi_k \int_c \frac{dz}{2\pi i} n_F(z) \frac{1}{z + i\omega_n - \xi_k + \frac{i}{2\tau_{z+m}}} \frac{1}{z - \xi_k + \frac{i}{2\tau_z}} \end{aligned} \quad (1.52)$$

The integrand is non-analytic if z or $z + i\omega_m$ is real so the integral is equal to the contributions from the contour as it passes just above and just below these lines. Let us assume that $\omega_n > 0$.

$$\begin{aligned}
& \langle j(0, \omega_n) j(0, -\omega_n) \rangle = \\
& -\frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi i} n_F(\epsilon) \frac{1}{\epsilon + i\omega_n - \xi_k + \frac{i}{2\tau}} \left(\frac{1}{\epsilon - \xi_k + \frac{i}{2\tau}} - \frac{1}{\epsilon - \xi_k - \frac{i}{2\tau}} \right) \\
& -\frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi i} n_F(\epsilon) \left(\frac{1}{\epsilon - \xi_k + \frac{i}{2\tau}} - \frac{1}{\epsilon - \xi_k - \frac{i}{2\tau}} \right) \frac{1}{\epsilon - i\omega_n - \xi_k - \frac{i}{2\tau}}
\end{aligned} \tag{1.53}$$

Only the second term in parenthesis in the first line of (1.53) contributes to the imaginary part of the integral and only the first term in parenthesis in the second line of (1.53) contributes to the imaginary part of the integral because the other terms have poles on the same side of the axis. Hence, we have

$$\begin{aligned}
& \langle j(0, \omega_n) j(0, -\omega_n) \rangle = \\
& \frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi i} n_F(\epsilon) \frac{1}{\epsilon + i\omega_n - \xi_k + \frac{i}{2\tau}} \frac{1}{\epsilon - \xi_k - \frac{i}{2\tau}} \\
& - \frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi i} n_F(\epsilon) \frac{1}{\epsilon - \xi_k + \frac{i}{2\tau}} \frac{1}{\epsilon - i\omega_n - \xi_k - \frac{i}{2\tau}}
\end{aligned} \tag{1.54}$$

Taking $i\omega_n \rightarrow \omega$ and dividing by ω , we have

$$\begin{aligned}
& \sigma(\omega) = \\
& \frac{1}{\omega} \frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} n_F(\epsilon) \frac{1}{\epsilon + \omega - \xi_k + \frac{i}{2\tau}} \frac{1}{\epsilon - \xi_k - \frac{i}{2\tau}} \\
& - \frac{1}{\omega} \frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} n_F(\epsilon) \frac{1}{\epsilon - \xi_k + \frac{i}{2\tau}} \frac{1}{\epsilon - \omega - \xi_k - \frac{i}{2\tau}}
\end{aligned} \tag{1.55}$$

Shifting $\epsilon \rightarrow \epsilon + \omega$ in the second integral and combining the two terms, we have:

$$\begin{aligned} \sigma(\omega) = & \\ \frac{1}{\omega} \frac{v_F^2}{3} N_F \int d\xi_k \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} (n_F(\epsilon) - n_F(\epsilon + \omega)) & \frac{1}{\epsilon + \omega - \xi_k + \frac{i}{2\tau}} \frac{1}{\epsilon - \xi_k - \frac{i}{2\tau}} \end{aligned} \quad (1.56)$$

Taking $\omega \rightarrow 0$ and noting that $(n_F(\epsilon) - n_F(\epsilon + \omega)) / \omega \rightarrow \delta(\epsilon)$, we have

$$\begin{aligned} \sigma_{\text{DC}} &= \frac{v_F^2}{6\pi} N_F \int d\xi_k \frac{1}{\xi_k^2 + (\frac{1}{2\tau})^2} \\ &= \frac{1}{3} v_F^2 \tau N_F \\ &\equiv \frac{1}{3\pi^2} k_F^2 \ell \end{aligned} \quad (1.57)$$

As expected, it is a constant determined by the lifetime due to impurity scattering. In the final line, we have written $v_F N_F = k_F^2 / \pi^2$. (This is the 3D expression. In general, it is proportional to k_F^{d-1} .)

The finite conductivity of electrons reflects the fact that electrons move diffusively rather than ballistically. If we define a diffusion constant $D = \frac{1}{3} v_F^2 \tau$, then $\sigma = N_F D$, which is the Einstein relation. Further insight into this can be gained by computing the density-density correlation function, to which we turn in the next section.

Note that we have not included vertex corrections in our calculation. This is not a problem because there are no vertex corrections to the *current vertex* to this order, as may be seen from the Ward identity, which relates vertex corrections to the derivative of the self-energy.

$$\Gamma^\mu(p, p, 0) = \frac{\partial \Sigma(p)}{\partial p_\mu} \quad (1.58)$$

As usual, we have used the notation $p_\mu = (\omega, p_i)$, $i = 1, 2, 3$. Since the self-energy is momentum-independent, there are no vertex corrections to the current vertices Γ^0 . On the other hand, the self-energy is energy-dependent, so there are important vertex corrections to the *density vertex*, Γ^0 .

Note that we computed this integral by performing the energy integrals first and then the momentum integrals. When these integrations do not commute, e.g. when the integrands are formally divergent by power-counting,

it is safest to introduce a momentum cutoff. When this is done, the order of integrations does not matter. A contour integration over ξ_k cannot be done, however, as a result of the cutoff, so the most natural way to proceed is to perform the energy integral first. The resulting momentum integral is then convergent in most cases. When the integrals are convergent, it doesn't matter in what order they are done. When they do not converge, as in the present case, the order does matter. The order in which we did the integrals is safest since it can be done with a momentum cutoff. If we had ignored this subtlety and done the integrals in the opposite order, and done the momentum integral without a cutoff (e.g. by contour integration) then we would have actually obtained the same result for the *real part* of the current-current correlation function. However, the imaginary part, which includes a term which cancels the diamagnetic term, would be missing.

1.5.4 Diffusion

The finite diffusion constant of electrons, D , which we found indirectly in the DC conductivity above, can be obtained explicitly by computing the density-density correlation function. As we will see, this correlation function has a very different form in a disordered electron system.

Consider the set of diagrams in figure (1.6)a (with, again, the electron Green function dressed with the one-loop self-energy, i.e. the lifetime). All of these diagrams contribute to the density-density correlation function. They form a geometric series which we can sum once we have obtained the value of the first. The higher-order diagrams constitute vertex corrections to the first diagram in the series. As noted in the previous subsection, there are no vertex corrections to the conductivity (i.e. the corresponding diagrams for the current-current correlation function do not correct the vertex in the DC limit, which is why we computed only a single diagram). However, such vertex corrections are important for the density-density correlation function. We will show that the resulting correlation function is consistent with the Ward identity.

The sum of these diagrams is

$$\langle \rho(q, \omega_m) \rho(-q, -\omega_m) \rangle = \frac{1}{\beta} \sum_n \int \frac{d^3\mathbf{k}}{(2\pi)^3} \Gamma(q, \epsilon_n, \epsilon_n + \omega_m) \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k} + \mathbf{q}, \epsilon_n + \omega_m) \quad (1.59)$$

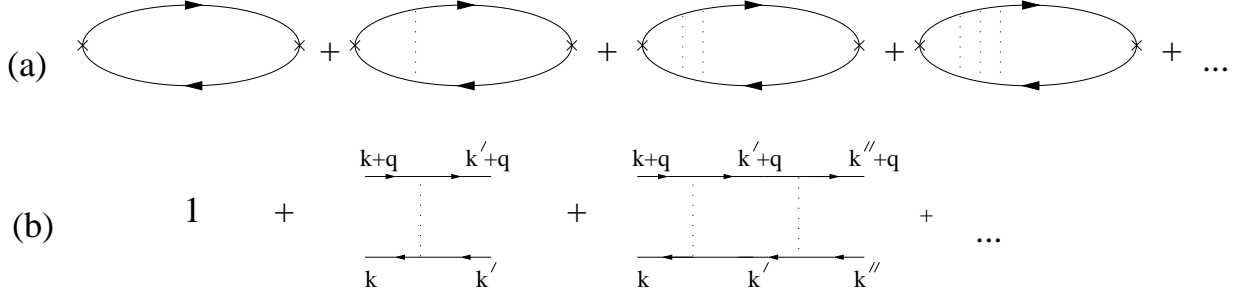


Figure 1.6: (a) A geometric series of diagrams which contribute to the density-density correlation function. (b) The quantity $\Gamma(q, \epsilon_n, \epsilon_n + \omega_m)$ which is the ladder sum which sits inside the density-density bubble.

where $\Gamma(q, \epsilon_n, \epsilon_n + \omega_m)$ is the infinite series of ladder diagrams in (1.6)b

$$\Gamma(q, \epsilon_n, \epsilon_n + \omega_m) = 1 + n_i v^2 I(q, \epsilon_n, \epsilon_n + \omega_m) + (n_i v^2)^2 (I(q, \epsilon_n, \epsilon_n + \omega_m))^2 + \dots \quad (1.60)$$

and the integral $I(q, \epsilon_n, \epsilon_n + \omega_m)$ is given by

$$I(q, \epsilon_n, \epsilon_n + \omega_m) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k} + \mathbf{q}, \epsilon_n + \omega_m) \quad (1.61)$$

Written explicitly,

$$I(q, \epsilon_n, \epsilon_n + \omega_m) = \frac{N_F}{2} \int d\xi_k d(\cos \theta) \frac{1}{i\epsilon_n + i\omega_m - \xi_k - v_F q \cos \theta + \frac{i}{2\tau_{n+m}}} \frac{1}{i\epsilon_n - \xi_k + \frac{i}{2\tau_n}} \quad (1.62)$$

where we have used the shorthand $\tau_n = \tau \text{sgn}(\epsilon_n)$. Integrating ξ_k , we have:

$$I(q, \epsilon_n, \epsilon_n + \omega_m) = i\pi N_F \int d(\cos \theta) \frac{\theta(\epsilon_n) - \theta(\epsilon_n + \omega_m)}{i\omega_m - v_F q \cos \theta + \frac{i}{\tau} [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]} \quad (1.63)$$

Integrating $\cos \theta$,

$$\begin{aligned}
I(q, \epsilon_n, \epsilon_n + \omega_m) &= \\
& \frac{i\pi N_F}{v_F q} [\theta(\epsilon_n) - \theta(\epsilon_n + \omega_m)] \ln \left(\frac{i\omega_m - v_F q + \frac{i}{\tau} [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]}{i\omega_m + v_F q + \frac{i}{\tau} [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]} \right) \\
&= \frac{i\pi N_F}{v_F q} [\theta(\epsilon_n) - \theta(\epsilon_n + \omega_m)] \ln \left(\frac{1 + (\omega_m \tau + i v_F \tau q) [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]}{1 + (\omega_m \tau - i v_F \tau q) [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]} \right)
\end{aligned} \tag{1.64}$$

Expanding to lowest non-trivial order in ω_m and q , we have

$$\begin{aligned}
I(q, \epsilon_n, \epsilon_n + \omega_m) &= -\frac{i\pi N_F}{v_F q} [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]^2 \times \\
& \left(2i v_F \tau q - 2i v_F \tau q \omega_m \tau [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)] + \frac{2}{3} (i v_F \tau q)^3 \right)
\end{aligned} \tag{1.65}$$

This expression vanishes unless $\epsilon_n, \epsilon_{n+m}$ have opposite signs, in which case:

$$\begin{aligned}
I(q, \epsilon_n(\epsilon_n + \omega_m) < 0) &= 2\pi N_F \tau \left(1 - |\omega_m| \tau - \frac{1}{3} v_F^2 \tau^2 q^2 \right) \\
&= 2\pi N_F \tau (1 - |\omega_m| \tau - D \tau q^2)
\end{aligned} \tag{1.66}$$

where the diffusion constant D is given by $D = v_F^2 \tau / 3$.

Summing the geometric series (1.60),

$$\begin{aligned}
\Gamma(q, \epsilon_n, \epsilon_n + \omega_m) &= \frac{1}{1 - n_i v^2 I(q, \epsilon_n, \epsilon_n + \omega_m)} \\
&= \frac{1}{1 - (1 - |\omega_m| \tau - D \tau q^2) [\theta(\epsilon_n + \omega_m) - \theta(\epsilon_n)]^2} \\
&= \frac{1}{\tau} \frac{1}{|\omega_m| + D q^2} \theta(-\epsilon_n \epsilon_{n+m}) + \theta(\epsilon_n \epsilon_{n+m})
\end{aligned} \tag{1.67}$$

where we have used $2\pi N_F \tau n_i v^2 = 1$. The same form is obtained in any dimension, with minor differences, such as $D = v_F^2 \tau / d$ in d dimensions. This form indicates that the electron density diffuses.

To make this a little more concrete, consider the density-density correlation function (1.59):

$$\begin{aligned}
\langle \rho(q, \omega_m) \rho(-q, -\omega_m) \rangle &= \frac{1}{\beta} \sum_n \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \theta(\epsilon_n \epsilon_{n+m}) \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k} + \mathbf{q}, \epsilon_n + \omega_m) \\
&+ \frac{1}{\beta} \sum_n \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{1}{\tau} \frac{1}{|\omega_m| + D q^2} \theta(-\epsilon_n \epsilon_{n+m}) \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k} + \mathbf{q}, \epsilon_n + \omega_m)
\end{aligned} \tag{1.68}$$

The first term is essentially the static response of free fermions with $\tau = \infty$ up to corrections of order $1/\tau$, ω_m :

$$\begin{aligned} \frac{1}{\beta} \sum_n \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \theta(\epsilon_n \epsilon_{n+m}) \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k} + \mathbf{q}, \epsilon_n + \omega_m) = \\ \frac{1}{\beta} \sum_n \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{1}{i\epsilon_n - \xi_k - v_F q \cos \theta} \frac{1}{i\epsilon_n - \xi_k} + O(1/\tau, \omega_m) \\ = -N_F + O(1/\tau, \omega_m) \quad (1.69) \end{aligned}$$

In order to do this integral correctly, we must integrate frequency first or else work with a cutoff in momentum space. Integrating in the opposite order will give us an answer which vanishes in the $\omega_m = 0$ limit, i.e. we will miss the static part. By separating the density-density correlation function into static and dynamic pieces – the two terms of (1.68) – we are left with a dynamic piece which is strongly convergent because the Matsubara sum is over a finite range. Hence, we can do the integrals in the opposite order in evaluating that term, to which we now turn.

The second term can be evaluated by doing a ξ_k contour integral:

$$\begin{aligned} \frac{1}{\beta} \sum_n \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{1}{\tau} \frac{1}{|\omega_m| + Dq^2} \theta(-\epsilon_n \epsilon_{n+m}) \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k} + \mathbf{q}, \epsilon_n + \omega_m) = \\ \frac{2\pi i}{\beta} \frac{N_F}{2} \sum_n \frac{1}{\tau} \frac{1}{|\omega_m| + Dq^2} \frac{[\theta(\epsilon_{n+m}) - \theta(\epsilon_n)]}{i\omega_m - v_F q \cos \theta + \frac{i}{\tau} [\theta(\epsilon_{n+m}) - \theta(\epsilon_n)]} \\ \approx \frac{2\pi i}{\beta} \frac{N_F}{2} \sum_n \frac{1}{\tau} \frac{1}{|\omega_m| + Dq^2} \frac{\theta(-\epsilon_n \epsilon_{n+m})}{\frac{i}{\tau}} \\ = N_F \frac{|\omega_m|}{|\omega_m| + Dq^2} \quad (1.70) \end{aligned}$$

Combining the two terms, we have

$$\langle \rho(q, \omega_m) \rho(-q, -\omega_m) \rangle = -N_F \frac{Dq^2}{|\omega_m| + Dq^2} \quad (1.71)$$

This correlation function has a pole at $|\omega_m| = -Dq^2$. In order to continue to real time, observe that $f(z) = -iz \operatorname{sgn}(\operatorname{Im}(z)) = |\omega_m|$ if $z = i\omega_m$. If we continue to $z = \omega + i\delta$, then we have $f(\omega + i\delta) = -i\omega$, so that the retarded density-density correlation function becomes:

$$\langle \rho(q, \omega) \rho(-q, -\omega) \rangle = -N_F \left(\frac{Dq^2}{-i\omega + Dq^2} \right) \quad (1.72)$$

The existence of a pole at $i\omega = -Dq^2$ means that the Fourier transform has the form:

$$\langle [\rho(\mathbf{x}, t), \rho(0, 0)] \rangle \theta(t) \sim \frac{1}{t^{d/2+1}} e^{-\frac{x^2}{2Dt}} \theta(t) \quad (1.73)$$

Hence, the electron density moves diffusively since a small, highly localized wavepacket prepared at $t = 0$ will spread out over a region of size $x \sim t^{1/2}$ at time t .

We obtained a diffusive form for the propagation of electron density by summing a particular class of diagrams, but these are clearly not the only possible diagrams. However, since the density-density correlation function which we obtained vanishes at $q = 0$, as required by the Ward identity, our neglect of other diagrams is at least consistent with charge conservation. There is a further justification for neglecting other diagrams: they are suppressed by a power of $1/k_F\ell$. However, we will find in the next section that such an argument is too quick.

With the density-density correlation function (1.72) in hand, we can re-visit the conductivity. Charge conservation tells us that

$$\partial_t \rho + \nabla \cdot \mathbf{j} = 0 \quad (1.74)$$

Fourier transforming this equation, we have the following relation between correlation functions:

$$q^2 \langle j(q, \omega) j(-q, -\omega) \rangle = \omega^2 \langle \rho(q, \omega) \rho(-q, -\omega) \rangle \quad (1.75)$$

Hence, from the previous expression for the density-density correlation function, we have

$$\frac{1}{i\omega} \langle j(q, \omega) j(-q, -\omega) \rangle = N_F \frac{i\omega D}{i\omega + Dq^2} \quad (1.76)$$

Taking $q = 0$, and then $\omega \rightarrow 0$, we have the DC conductivity.

The spin-spin correlation function also has a diffusive form. Consider, for instance, the $S_z - S_z$ correlation function. For non-interacting electrons, it is precisely the same as the density-density correlation function because $S_z = \rho_\uparrow - \rho_\downarrow$ implies that $\langle S_z S_z \rangle = \langle \rho_\uparrow \rho_\uparrow \rangle + \langle \rho_\downarrow \rho_\downarrow \rangle = \langle \rho \rho \rangle$ (the cross-terms vanish for non-interacting electrons). by rotational symmetry, the other spin-spin correlation functions must be identical. This can also be easily seen by direct calculation of $\langle S_+ S_- \rangle$: the electron line will have spin $+1/2$ while the hole line will have spin $-1/2$, but the diagrams will be unchanged from the density-density calculation because the fermion propagators are spin-independent:

$$\langle S_+(q, \omega) S_-(-q, -\omega) \rangle = -N_F \left(\frac{Dq^2}{-i\omega + Dq^2} \right) \quad (1.77)$$

In the presence of impurities, electrons no longer move ballistically at long length scales. They move diffusively because they undergo many collisions with impurities. We can think of this in the following RG language. Consider the effective Lagrangian for a Fermi liquid in the presence of impurities.

$$S = \frac{k_F^{d-1}}{(2\pi)^d} \int dl d^d\Omega \frac{d\epsilon}{2\pi} \psi^\dagger (i\epsilon - v_F l) \psi + \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{d^3\mathbf{k}'}{(2\pi)^3} \frac{d\epsilon}{2\pi} V(\mathbf{k} - \mathbf{k}') \psi^\dagger(\mathbf{k}, \epsilon) \psi(\mathbf{k}', \epsilon) \quad (1.78)$$

To avoid clutter, we suppress the marginal four-Fermi interactions parametrized by Landau parameters which are inessential to this discussion.

For a fixed $V(\mathbf{k})$, this term is precisely marginal. This is what would be expected from a single-impurity. Consider, now, a random $V(\mathbf{k})$. When we integrate over different impurity configurations according to (1.46) we generate the effective action

$$S_{\text{eff}} = \frac{k_F^{d-1}}{(2\pi)^d} \int dl d^d\Omega \frac{d\epsilon}{2\pi} \psi^\dagger (i\epsilon - v_F l) \psi + n_i v^2 \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{d^3\mathbf{k}'}{(2\pi)^3} \frac{d\epsilon}{2\pi} \frac{d^3\mathbf{p}}{(2\pi)^3} \frac{d^3\mathbf{p}'}{(2\pi)^3} \frac{d\epsilon'}{2\pi} \psi^\dagger(\mathbf{k}, \epsilon) \psi(\mathbf{k}', \epsilon) \psi^\dagger(\mathbf{p}, \epsilon') \psi(\mathbf{p}', \epsilon') \quad (1.79)$$

with the proviso that we should only allow diagrams which would remain connected even if all of the impurity lines (connecting the two pairs of fermions in the second line in equation (1.79)) were cut. (We will introduce a formal way of implementing this in the next chapter.) The resulting impurity-scattering term is strongly-relevant, scaling as s^{-1} . (It is similar to a four-Fermi interaction but with one fewer energy integration.) Thus the clean Fermi liquid is unstable to impurity-scattering. The RG flow is towards the *diffusive Fermi liquid*. At this fixed point, the density-density and spin-spin correlation functions takes the diffusive form (1.72), (1.77) while the single-fermion Green function (1.48) is short-ranged as a result of the i/τ in the denominator. In the next chapter, we will construct an effective field theory for the long-wavelength, low frequency modes of the diffusive fixed point.

1.5.5 Weak Localization

In the previous two subsections, we expended substantial effort to compute the electron Green function, and the density-density and spin-spin correla-

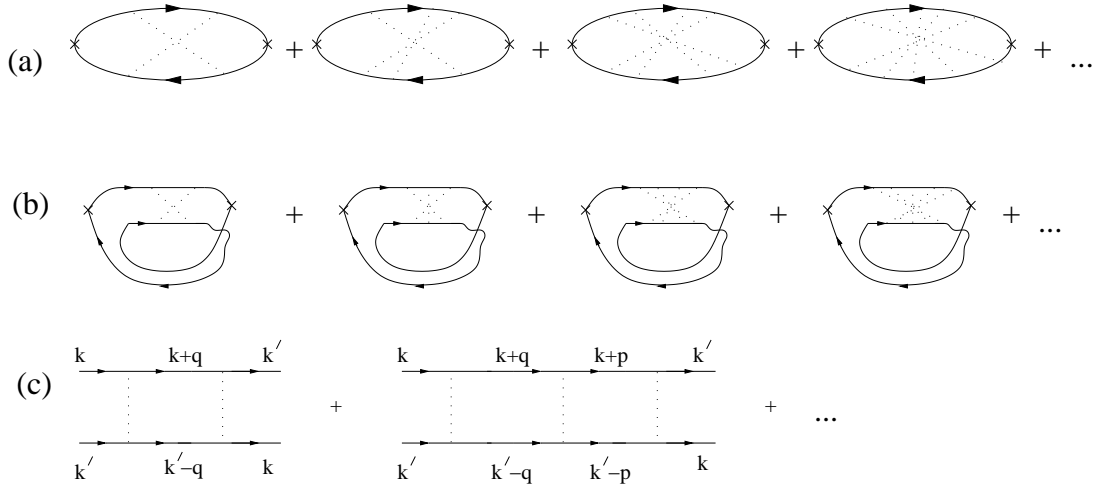


Figure 1.7: (a) Diagrams contributing to the conductivity with impurity lines maximally-crossed. (b) A re-drawing of the maximally-crossed diagrams which emphasizes the role of the particle-particle diagrams shown in (c).

tion functions. However, the result of all of these diagrammatic calculations was no more than simple qualitative considerations and naive Drude theory would have given us: electrons have a finite conductivity which is determined by the lifetime τ . This lifetime also enters the diffusive form of various correlation functions through the diffusion constant $D = v_F^2 \tau / d$. One hardly needs sophisticated field-theoretic techniques to discover this, although it is, perhaps, some consolation that we can confirm our intuition by direct calculation.

One might naively think that, in the absence of interactions, this is all that there is to the conductivity of a dirty metal. However, the sub-leading impurity contributions to the conductivity are also interesting. They are due to quantum interference effects. In two dimensions, the ‘subleading’ correction is divergent at low-temperatures and, in fact, non-interacting electrons are always insulating in $d = 2$. Field-theoretic techniques are invaluable in obtaining and understanding these corrections.

Consider the diagrams of figure 1.7a. In these diagrams, the impurity lines are maximally-crossed. This is a particular class of diagrams contributing to the conductivity. In this subsection, we will see why these diagrams give a significant contribution to the conductivity. Observe that they can

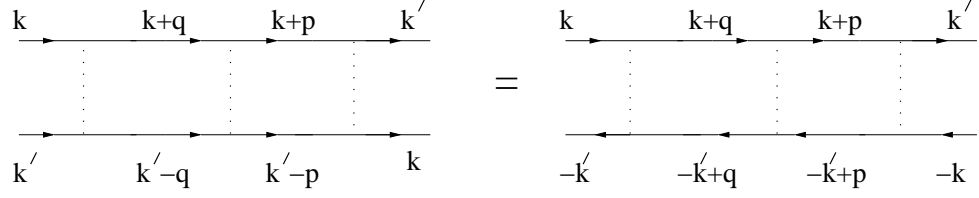


Figure 1.8: A particle-particle diagram is equal to a particle-hole diagram with particle momenta reversed to become hole momenta.

be re-drawn in the manner shown in figure 1.7b. The diagrams of 1.7b are very similar to those of the diffusion propagator. The main difference is that the arrows are in the same direction, rather than opposite directions, i.e. the middle of the diagram is a particle-particle diagram rather than a particle-hole diagram. However, time-reversal symmetry relates these two. Call the sum of particle-particle diagrams in figure 1.7c $W(\mathbf{k}, \mathbf{k}', \epsilon_n, \epsilon_{n+m})$. By time-reversal symmetry, we can reverse the direction of one of the arrows if we also reverse the sign of the momentum. Because particles and holes both cost positive energy, but carry opposite momentum, a hole line running through a diagram can be replaced with a particle line at the same energies but opposite momenta without changing the value of the diagram. This is depicted in figure 1.8. From this observation, we see that:

$$\begin{aligned}
 W(\mathbf{k}, \mathbf{k}', \epsilon_n, \epsilon_{n+m}) &= (n_i v^2)^2 I(\mathbf{k} + \mathbf{k}', \epsilon_n, \epsilon_{n+m}) \\
 &\quad + (n_i v^2)^3 (I(\mathbf{k} + \mathbf{k}', \epsilon_n, \epsilon_{n+m}))^2 + \dots \\
 &= \frac{(n_i v^2)^2 I(\mathbf{k} + \mathbf{k}', \epsilon_n, \epsilon_{n+m})}{1 - (n_i v^2) I(\mathbf{k} + \mathbf{k}', \epsilon_n, \epsilon_{n+m})} \\
 &= \frac{n_i v^2}{\tau} \left(\frac{1}{|\omega_m| + D(\mathbf{k} + \mathbf{k}')^2} \right) \theta(-\epsilon_n \epsilon_{n+m}) \quad (1.80)
 \end{aligned}$$

If we substitute this into a conductivity bubble to compute its contribution to the conductivity, as in figure 1.7b, we obtain

$$\begin{aligned}
 \delta\sigma(\omega_m) &= \frac{1}{\omega_m} \frac{v_F^2}{d} \int \frac{d^d k}{(2\pi)^d} \frac{d^d k'}{(2\pi)^d} \frac{1}{\beta} \sum_n \mathcal{G}(\mathbf{k}, \epsilon_n) \mathcal{G}(\mathbf{k}, \epsilon_n + \omega_m) \times \\
 &\quad W(\mathbf{k}, \mathbf{k}', \epsilon_n, \epsilon_{n+m}) \mathcal{G}(\mathbf{k}', \epsilon_n) \mathcal{G}(\mathbf{k}', \epsilon_n + \omega_m) \quad (1.81)
 \end{aligned}$$

From (1.80), we see that $W(\mathbf{k}, \mathbf{k}', \epsilon_n, \epsilon_{n+m})$ is sharply peaked at $\mathbf{k} \approx -\mathbf{k}'$. Hence, we write $\mathbf{q} = \mathbf{k} + \mathbf{k}'$ and change the variables of integration to \mathbf{k}, \mathbf{q} .

In the second pair of Green functions, we take $\mathbf{k} = -\mathbf{k}'$, neglecting the weak \mathbf{q} dependence of those factors. We now have:

$$\begin{aligned}
\delta\sigma(\omega_m) &= \frac{1}{\omega_m} N_F \frac{v_F^2}{d} \int d\xi_k \int \frac{d^d q}{(2\pi)^d} \frac{1}{\beta} \sum_n \frac{n_i v^2}{\tau} \left(\frac{\theta(-\epsilon_n \epsilon_{n+m})}{|\omega_m| + Dq^2} \right) \times \\
&\quad \left(\frac{1}{i\epsilon_n - \xi_k + \frac{i}{2\tau_n}} \right)^2 \left(\frac{1}{i\epsilon_n + i\omega_m - \xi_k + \frac{i}{2\tau_{n+m}}} \right)^2 \\
&= \frac{4\pi i}{\omega_m} N_F \frac{v_F^2}{d} \int \frac{d^d q}{(2\pi)^d} \frac{1}{\beta} \sum_n \frac{n_i v^2}{\tau} \left(\frac{\theta(\epsilon_{n+m}) - \theta(\epsilon_n)}{|\omega_m| + Dq^2} \right) \times \\
&\quad \left(\frac{1}{i\omega_m + \frac{i}{\tau}(\theta(\epsilon_{n+m}) - \theta(\epsilon_n))} \right)^3 \\
&\approx \frac{4\pi i}{\omega_m} N_F \frac{v_F^2}{d} \int \frac{d^d q}{(2\pi)^d} \frac{1}{\beta} \sum_n \frac{n_i v^2}{\tau} \left(\frac{\theta(\epsilon_{n+m}) - \theta(\epsilon_n)}{|\omega_m| + Dq^2} \right) \times \\
&\quad \left(\frac{1}{\frac{i}{\tau}(\theta(\epsilon_{n+m}) - \theta(\epsilon_n))} \right)^3 \\
&= -\frac{4\pi}{\omega_m} N_F \frac{v_F^2}{d} \tau^3 \int \frac{d^d q}{(2\pi)^d} \frac{1}{\beta} \sum_n \frac{n_i v^2}{\tau} \left(\frac{\theta(-\epsilon_{n+m} \epsilon_n)}{|\omega_m| + Dq^2} \right) \\
&= -\frac{1}{\pi} D \int \frac{d^d q}{(2\pi)^d} \frac{1}{|\omega_m| + Dq^2} \\
&= -\frac{1}{\pi} \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \frac{1}{(2\pi)^d} \frac{1}{d-2} \left(\frac{1}{\ell^{d-2}} - \left(\frac{|\omega_m|}{D} \right)^{\frac{d-2}{2}} \right) \quad (1.82)
\end{aligned}$$

We have taken the ultraviolet cutoff to be $1/\ell$. For shorter wavelengths, one should use ballistic, rather than diffusive propagators.

While (1.81) is a small correction in the low frequency limit for $d > 2$, it is divergent in this limit for $d < 2$. In $d = 2$, it is;

$$\begin{aligned}
\delta\sigma(\omega_m) &= -\frac{1}{4\pi^2} \ln \left(\frac{D}{|\omega_m| \ell^2} \right) \\
\frac{\delta\sigma(\omega_m)}{\sigma_0} &= -\frac{1}{\pi k_F \ell} \ln \left(\frac{D}{|\omega_m| \ell^2} \right) \quad (1.83)
\end{aligned}$$

which is also divergent.

Hence, in $d \leq 2$, the correction (1.81)-(1.83), which is formally smaller than the (semiclassical) Boltzmann result by a power of $k_F \ell \sim D$ is, in

fact divergent as $\omega \rightarrow 0$. Thus, the starting point of the diffusive fixed point is unstable. In the next section, we will discuss the result of this instability in more detail but, for now, we observe that it suppresses the conductivity. Thus, it drives the system towards localization. In $d = 2$, when this correction is still weak (sufficiently high frequencies or, as we will describe below, temperatures) and growing only logarithmically, it is called *weak localization*.

One might wonder whether it is valid to stop with the diagrams which we have calculated. Isn't it possible that if we computed some more diagrams, we would find other divergent corrections to the conductivity which might change our conclusions entirely? The answer is that this will not happen and we can stop here. The divergences which we find are due to the existence of slow modes in the system. The diffusive mode is a slow mode because it is related to the conservation of charge. The particle-particle mode (called a 'Cooperon') is slow because it is related to the diffusion mode by time-reversal symmetry. There is no other mode which is guaranteed to be slow by symmetry, so we do not need to worry about other entirely new divergences, although these same slow modes can lead to subleading divergences. We will discuss this point in the next chapter using an effective field theory for the slow modes of the system.

Thus far, we have considered non-interacting electrons and neglected phonons, etc. If these sources of *inelastic scattering* are considered, they cut off the above divergences since they break the symmetry which allowed us to substitute a hole line for a particle line. For length scales longer than the Thouless length

$$\ell_{\text{Thouless}} = \sqrt{D\tau_{\text{in}}} \quad (1.84)$$

inelastic processes will thwart localization and cause ohmic conduction. For example, the $2D$ weak-localization correction is:

$$\delta\sigma(\omega_m) = -\frac{1}{2\pi N_F \tau^2} \frac{1}{2\pi D} \ln(\ell_{\text{Thouless}}/\ell) \quad (1.85)$$

If $\tau_{\text{in}} \sim T^{-p}$, then the frequency is replaced by $\sim T^p$, *not* T , as one might naively assume.

A weak magnetic field will also cut off localization at length scales longer than the magnetic length $\ell_H = 1/\sqrt{H}$. (Restoring the fundamental constants, it is $\ell_H = \sqrt{\hbar c/eH}$.) Thus, in the weak localization regime, where

$$\delta\sigma(\omega_m) = -\frac{1}{2\pi N_F \tau^2} \frac{1}{2\pi D} \ln(\ell_H/\ell) \quad (1.86)$$

the conductivity *increases* as the magnetic field is increased, i.e. the magnetoresistance is *negative*.

1.5.6 Quasiclassical Approach to Weak Localization***

1.5.7 Weak Magnetic Fields and Spin-orbit Interactions: the Unitary and Symplectic Ensembles

1.6 The Metal-Insulator Transition

1.6.1 Percolation

Consider a disordered solid which is insulating as a result of impurities. Although it is insulating, there will be some isolated regions which are ‘conducting’. In a system of non-interacting fermions, such regions would have single-particle states at the Fermi energy which extend across them but do not leak substantially into the surrounding insulating regions. One way to think about this is to consider the potential $V(x)$, whose spatial variation is random as a result of impurities. The regions in which $V(x) < \epsilon_F$ are classically-allowed for electrons at the Fermi energy. Let us consider these to be the ‘conducting’ regions. The regions in which $V(x) > \epsilon_F$ are classically-forbidden for electrons at (or below) the Fermi energy. Let us call these ‘insulating’ regions.

Suppose we decrease the impurity concentration. Then the conducting regions will grow in size. As we continue to decrease the impurity concentration, the conducting regions will grow larger and larger and will begin to merge. Eventually, it will become possible to go from one end of the system to the other within one long conducting region which spans the system. If electrical conduction were purely classical and ohmic, this point – which is called the *percolation threshold* – would be the transition point between metallic and insulating behavior.

However, electrical conduction is not classical. Even below the percolation threshold, electrons can tunnel quantum mechanically from one conducting region to another. Conversely, an electron can take two different paths from point A to point B , and the quantum mechanical amplitudes for these processes might interfere destructively. Nevertheless, at temperatures which are higher than the characteristic phase coherence temperatures of the system – so that coherent quantum tunneling cannot occur – but not so high that the conductivity of the ‘insulating’ regions (insulating and metallic are precise distinctions only at zero temperature; at finite temperatures,

all systems have finite conductivity) is comparable to that of the conducting ones, we expect the conductivity and other physical properties to show behavior characteristic of *percolation* processes. Since percolation is, in a sense, the classical limit of the metal-insulator transition, we will discuss it briefly.

There are two basic lattice models of percolation, *site percolation* and *bond percolation*. In site percolation, one removes sites randomly from a lattice and considers the properties of the remaining sites. We will say that a site is ‘present’ if it has not been removed and ‘absent’ if it has been. We imagine that an electron can hop directly from a given site which is present to its nearest neighbors if they are present, but not to other sites. Thus, if there is an isolated present site which is surrounded by absent sites, then an electron on that site is stuck there. Sites which are present can be grouped into clusters of sites which are contiguous with each other. One site in a cluster can be reached from any other site in the cluster by a sequence of nearest-neighbor hops entirely within the cluster. In bond percolation, one removes bonds from a lattice. An electron can hop from one lattice site to a nearest-neighbor site only if the bond between them is present. Sites can again be grouped into clusters which are connected by bonds so that any site in cluster can be reached from any other site in the cluster.

For the sake of concreteness, let us consider *bond percolation*. When the average bond density, which we will call p , is low, clusters tend to be small. As the bond density is increased, clusters grow until the percolation threshold, p_c , is reached. At this point, one cluster extends across the length of the system; it is called the *percolating cluster*. For $p < p_c$, the average distance between two sites on the same cluster, ξ , scales as

$$\xi \sim (p_c - p)^{-\nu} \quad (1.87)$$

For $p > p_c$, the fraction, P , of the system which belongs to the percolating cluster is

$$P(p) \sim (p - p_c)^\beta \quad (1.88)$$

(For $p < p_c$, $P = 0$ because the largest cluster in an infinite system is finite.) Notice the analogy between these exponents and the correlation length and order-parameter exponents associated with critical phenomena.

One might imagine that the conductivity (assuming that bonds which are present are considered to be conducting and those which are absent are insulating) would be proportional to the fractions of the system belonging to the largest cluster. However, this is not the case because many of the

bonds in the cluster lead to dead-ends which do not contribute to the conductivity. Instead, the conductivity is associated with a different exponent, μ , (sometimes called t) which describes the ‘backbone’ of the percolating cluster – i.e. the cluster with dead-ends (those bonds which do not have two completely independent routes to the two ends of the cluster) removed.

$$\sigma(p) \sim (p - p_c)^\mu \quad (1.89)$$

At the percolation point itself, the system will have power-law correlations. Many of the properties of the percolation point in two dimensions have been revealed using conformal field theory and, more recently with the use of the stochastic Loewner equation.

However, as we pointed out earlier, percolation does not correctly describe metal-insulator transitions because it misses quantum-mechanical effects. In the next section, we discuss alternate approaches which take these effects into account.

1.6.2 Mobility Edge, Minimum Metallic Conductivity

Let us now consider a metal-insulator transition in a system of non-interacting fermions. Such a transition can occur as a result of varying either the chemical potential or the disorder strength. If the system is weakly-disordered, then the energy spectrum will still approximately break up into bands, although there may not be a true gap between the bands because there will be exponentially-small tails in the density of states. Near the bottom of a band, the states will all be localized because the kinetic energies of these states will be low. Near the top of the band, the same will be true. It is useful to think of a band with chemical potential near the top as a nearly empty hole band; in such a case, the hole kinetic energies near are low near the Fermi energy. Near the middle of the band, on the other hand, the states will be extended if the disorder strength is sufficiently small because the kinetic energies will be large in comparison. (If the disorder strength is too large, then even at the middle of the band the kinetic energy will not be large enough to overcome it.) What happens as the Fermi energy is swept from the bottom of the band to the middle? To answer this question, one should first note that there are no energies at which there are both extended and localized states because localization is not stable in the presence of extended states at the same energy. A small change in the particular realization of disorder (e.g. the locations of the impurities) would cause mixing between the localized states and the extended ones. When a localized state mixes with extended states, it becomes extended.

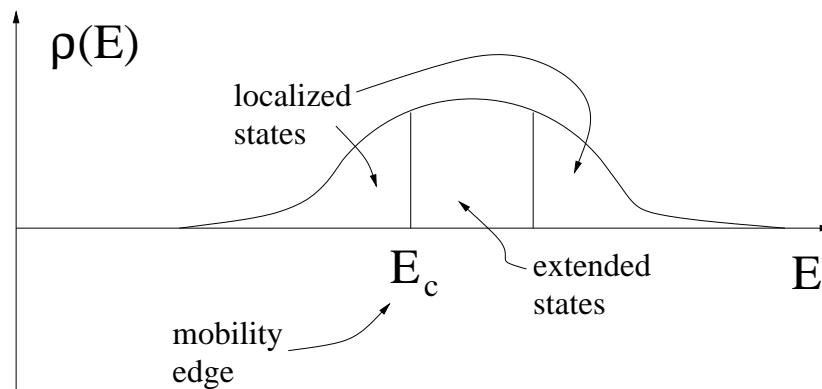


Figure 1.9: The density-of-states in a disordered system, with extended and localized states and mobility edges shown.

Hence, there will be sharp energies which separate localized states from extended ones. Such an energy is called a *mobility edge*. As depicted in figure 1.9, the mobility edge separates the strongly-localized states with low kinetic energy in the band tails from the extended states with high kinetic energy in the center of the band. (One can imagine there being more than a single pair of such energies, but this is not generic and would require some special circumstances.)

As the chemical potential is moved through the mobility edge, E_c , the system undergoes a metal-insulator transition. For $E_F < E_c$, the zero-temperature DC conductivity σ vanishes. For $E_F > E_c$, it is finite. What happens at $E_F = E_c$? Is it equal to some minimum non-zero value σ_{\min} , the *minimum metallic conductivity*? According to the percolation model described in the preceding section, the conductivity vanishes precisely at the transition, $\sigma \sim |E_F - E_c|^\mu$. ($E_F - E_c$ or any other control parameter will act as a proxy for $p - p_c$ near the transition.) On the other hand, the following heuristic argument suggests that it is finite. The conductivity is given in Boltzmann transport theory by

$$\sigma = A_d \frac{e^2}{\hbar} k_F^{d-1} \ell \quad (1.90)$$

where ℓ is the mean-free-path and A_d is a dimension-dependent constant. (Here, we have restored the factors of e , \hbar which were set to one earlier.) If scattering is due entirely to impurities, the mean-free path cannot be shorter than the inter-atomic spacing a in the solid. Hence, $k_F \ell \gtrsim \pi$ since $k_F \sim \pi/a$.

Hence, we conclude that

$$\sigma_{\min} = A_d \frac{e^2}{\hbar} k_F^{d-2} \quad (1.91)$$

In particular, in $d = 2$, the minimum metallic conductivity can be written entirely in terms of fundamental constants, $\sigma_{\min} = A_d e^2 / \hbar$ (restoring the fundamental constants which we have set to 1).

However, one should be cautious in applying the Boltzmann transport formula (1.90) – which is semiclassical and valid for large $k_F \ell$ – in the regime $k_F \ell \sim 1$. Indeed, as we will see in the next subsection, the preceding logic fails in the case of non-interacting electrons.

1.6.3 Scaling Theory

Thouless had the important insight that one should study the scaling behavior of the conductance. Consider a system of non-interacting electrons of size L . Thouless wrote the conductance of the system in the following form, which we encountered earlier (1.6):

$$g(L) = \frac{\overline{\delta E}}{\overline{\Delta E}} \quad (1.92)$$

δE is the disorder-averaged energy change of eigenstates at E_F when the boundary conditions are changed from periodic to antiperiodic. ΔE is the mean level spacing of eigenstates. (In order to derive such a relation analytically, one must consider disorder-averaged quantities.)

To understand this formula, suppose that $D(L)$ is the diffusion constant for an electron at the Fermi energy. The time required for an electron to diffuse across the system is the Thouless time:

$$t_L = \frac{L^2}{D(L)} \quad (1.93)$$

According to the uncertainty principle,

$$\overline{\delta E} \sim \frac{1}{t_L} \sim \frac{D(L)}{L^2} \quad (1.94)$$

Meanwhile,

$$\overline{\Delta E} = \frac{1}{N_F L^d} \quad (1.95)$$

Hence,

$$\frac{\overline{\delta E}}{\overline{\Delta E}} = N_F D(L) L^{d-2} = \sigma L^{d-2} = g(L) \quad (1.96)$$

The second equality follows from the Einstein relation between the diffusion constant and the conductivity.

We now study how $g(L)$ scales as the system size is increased. We could increase the system size from L to bL by stacking b subsystems of size L . The key assumption of Abrahams, Anderson, Licciardello, and Ramakrishnan ('the gang of four') is that:

$$g(bL) = f(b, g(L)) \quad (1.97)$$

In particular, L does not appear explicitly on the right-hand-side. In a macroscopic ohmic system, such a relation is obvious, $\frac{g(bL)}{g(L)} = b^{d-2}$. However, this is not exactly right microscopically. While $\frac{\Delta E(bL)}{\Delta E(L)} = b^{-d}$ should hold generally, we need to find a relation for $\frac{\delta E(bL)}{\delta E(L)}$. This relation might be complicated since we have to match the eigenfunctions from each size L block at their boundaries in order to construct the eigenfunctions of the size bL system. However, the sensitivity of the eigenfunctions to their boundary conditions is precisely what is encapsulated in $g(L)$. Thus, it is very reasonable to assume that $g(bL)$ depends only on b and $g(L)$. As noted above, this assumption holds in the ohmic regime. It also holds in the localized regime, where $g(L) = g_0 e^{-L/\xi}$ (ξ is the localization length), since $g(bL) = g_0 (g(L)/g_0)^b$. The scaling assumption of the gang of four was that the form (1.97) holds throughout the entire region between these limits.

If (1.97) holds, then

$$\frac{d \ln g}{d \ln L} = \frac{1}{g} \left(\frac{df}{db} \right)_{b=1} = \beta(g) \quad (1.98)$$

In strongly-disordered systems, g is small, and all states are localized. In this limit, we expect $\beta(g) = \ln(g/g_0)$. In the clean, or large g , limit, we expect Ohmic scaling, and $\beta(g) = d - 2$.

How do we connect the two limits? For g large, we can compute $\beta(g)$ in powers of $1/g$ or, equivalently, $1/k_F \ell$. In the previous section, we have already computed the first correction in $1/k_F \ell$ to the conductivity. If we replace the inelastic scattering time with the system size as an infrared cutoff, then this expression gives us $\sigma(L)$ for large $k_F \ell$. Differentiating, we obtain $\beta(g)$.

$$\beta(g) = d - 2 - \frac{a}{g} + \dots \quad (1.99)$$

where \dots represents $O(1/g^2)$ terms. From our computation of the weak-localization correction to the conductivity in $d = 2$, $a = 1/4\pi^2$. This is interpolated with the small g limit in figure 1.10.

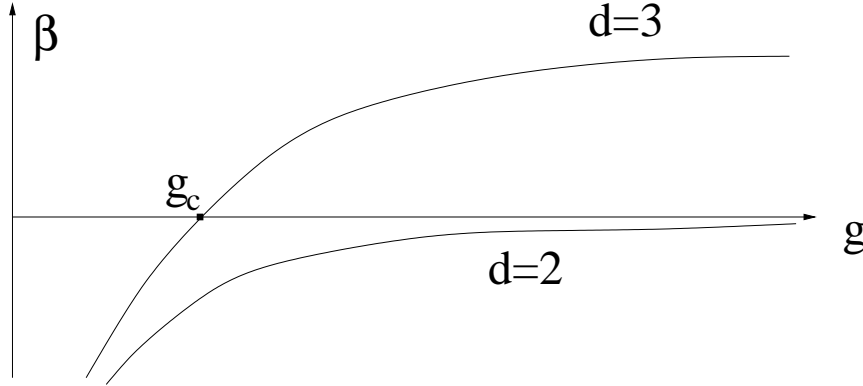


Figure 1.10: The β -function of Abrahams, *et al.* which is obtained by smoothly connecting the Ohmic limit with the localized one.

There are several consequences which follow immediately. (1) There is no metallic state of non-interacting fermions in $d \leq 2$ (and, therefore, no metal-insulator transition, of course). (2) There is a metal-insulator transition with no minimum metallic conductivity in $d > 2$. (3) For $d = 2 + \epsilon$, the metal-insulator transition occurs as $g \sim 1/\epsilon$, which is in the perturbative regime. The first statement is clear from figure 1.10: since $\beta(g) < 0$ for all g , the conductance always flows to the localized regime of small g . The second statement follows from the positivity of $\beta(g)$ for large g . In order to connect to the negative $\beta(g)$ localized regime at small g , $\beta(g)$ must pass through zero at some g_c . For $g > g_c$, $\beta(g) > 0$ and the conductance flows to large values, where it is Ohmic; this is the metallic state. For $g < g_c$, $\beta(g) < 0$ and the conductance flows to small values; this is the insulating state. For $g = g_c$, the conductance remains constant; this is the transition point. However, if the conductance remains constant as the system size L goes to infinity, the conductivity $\sigma_c = g_c L^{2-d} \rightarrow 0$. Hence, there is no minimum metallic conductivity.

The third statement follows immediately from (1.99). We can say more about the critical region by expanding the β -function about g_c :

$$\beta(g) = \frac{1}{\nu} \left(\frac{g - g_c}{g_c} \right) \quad (1.100)$$

From our β -function (1.99), we see that $1/\nu = \epsilon$ and $g_c = a_d/\epsilon$, where $\epsilon = d - 2$. Since the ... in (1.99) can be neglected only for g large, we can trust our $g_c = a_d/\epsilon$ conclusion only for ϵ small, i.e. only near two dimensions.

Suppose $|g - g_c|$ is small at the microscopic scale ℓ . The linearized β -function is applicable. Let us integrate until $|g - g_c|/g_c \sim 1$. Let us define the length scale at which this is reached as ξ :

$$\xi = \ell \left| \frac{g - g_c}{g_c} \right|^{-\nu} \quad (1.101)$$

This is an important length scale because the system will cross over to Ohmic metallic ($g < g_c$) or localized insulating ($g > g_c$) behavior at longer length scales. On the insulating side, ξ is the localization length.

On the conducting side, ξ is a classical-quantum crossover length. For length scales longer than ξ , electrical conduction is Ohmic and, therefore, essentially classical. For length scales shorter than ξ , quantum interference effects can occur and classical ideas are inappropriate. The conductivity is equal to

$$\sigma = \frac{g_c}{\xi} \quad (1.102)$$

The vanishing of the minimum metallic conductivity can be understood as the divergence of ξ at the metal-insulator transition.

Thus, the scaling hypothesis of Abrahams, *et. al* is equivalent to the assumption that there is a single length scale, ξ , which becomes large near the metal-insulator transition. This is completely analogous to the case of classical thermal phase transitions studied in chapter .. or quantum phase transitions studied in chapter .. However, there is an important difference: *a metal-insulator transition of non-interacting fermions is not a thermodynamic phase transition.* While the conductivity changes sharply at the transition, all thermodynamic properties, such as the ground state energy, are perfectly smooth across the transition. The smoothness of the ground state energy, for instance, follows from the smoothness of the density-of-states across the transition.