

Mesoscopics and Quantum Transport

Angus MacKinnon
Imperial College London

September 4, 2003

1 Introduction

Almost all the solid state physics you learnt as an undergraduate has been concerned with crystals, sometimes with a relatively low density of defects. The explanations of the various effects have made extensive use of Bloch's Theorem. Concepts such as Brillouin zones, bands, vertical transitions, effective mass and heavy and light holes are really only well defined in a perfect infinite crystal. In the absence of crystallinity none of these concepts is valid. A new approach is required.

When we consider either disordered or mesoscopic systems we are no longer dealing with infinite perfect crystals: disordered systems are not perfect and mesoscopic ones are not infinite. Thus Bloch's theorem and everything which derives from it can no longer be used as the basis for our description of the physics of the system. For several decades it was believed that the physics of non-crystalline solids, such as glasses, organic materials, alloys, etc., was basically the same as crystalline solids with some of the features, such as band edges or van Hove singularities washed out. We now know that this is not the case and that there is a number of effects, notably the metal-insulator transition which are peculiar to non-crystalline systems.

2 Localisation

2.1 Percolation

Let us start by considering a simpler classical problem. How does a fluid flow through a random medium? This is a problem of considerable importance in its own right: the extraction of oil from porous rock strata.

Consider a random landscape which is being slowly filled with water. At first (fig. 1 – left) there will be a continuous land mass with a few lakes. When the water level is very high (fig. 1 – right) we have islands in a sea. Let us now suppose there is a dam at the edge of the area which requires large quantities of water to drive a power station. When the water level is low only the lake next to the dam can be used and will soon run out. As the level is raised this lake becomes larger but still finite. The power station will run longer but will still eventually have to stop. Eventually at a critical water level (fig. 1 – centre) the system changes from a lake district to an archipelago. After this the power station can run indefinitely without fear of running out of water.

This phenomenon has much in common with more conventional phase transitions. The power station will stop running if the water freezes. There is a characteristic length scale which diverges at the transition: the size of the lakes or islands.

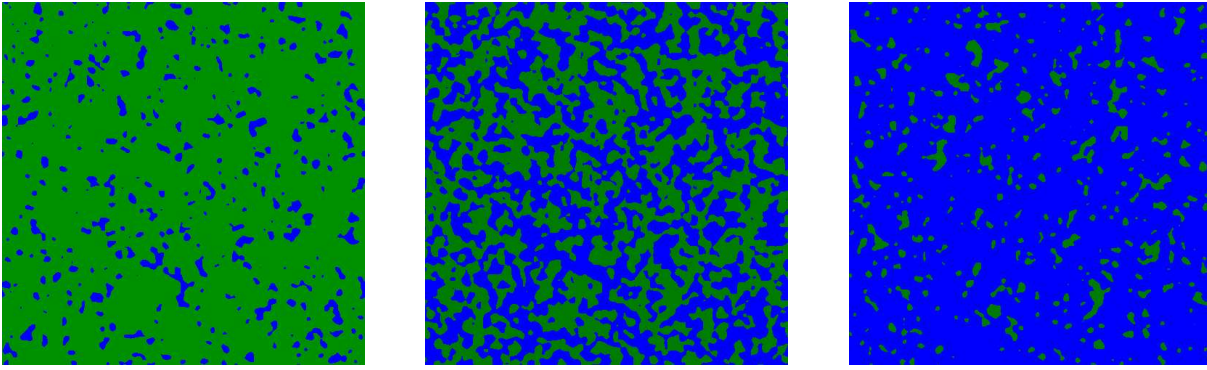


Figure 1: A random landscape filling with water. Left: low water level, isolated lakes. Centre: critical water level. Right: high water level, isolated islands.

There is a well defined critical water level rather like the critical temperature of the freezing transition or the ferromagnetic to paramagnetic transition in iron. If we think in terms of the density of blockages rather than the water level we see that there is a critical density above which the flow of water stops.

The one dimensional version of this problem is special. Any blockage of the channel is enough to prevent the flow of water. The critical density is zero. In higher dimensions, in contrast, water can flow round the blockage.

This is an example of a problem which cannot be solved by perturbation theory. There is a discontinuous jump in the behaviour between a system with no blockages and one with a single blockage.

Problem 1 Show that the critical water level for a random landscape occurs when the system consists of 50% land and 50% water? Consider the symmetry between water and land and the possibility of a boundary spanning the system in an isotropic system.

2.2 The Anderson Transition and the Mobility Edge



Figure 2: Phil Anderson

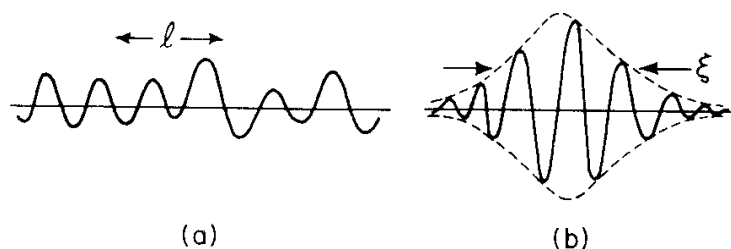


Figure 3: Schematic diagrams of (a) extended and (b) localised states, showing the correlation length, ℓ , and the localisation length, ξ .

Anderson (1958) (fig. 2) introduced the concept of localisation of electrons due to disorder. He argued that an electron which starts at a particular site cannot completely diffuse away from

that site if the disorder is greater than some critical value. He thus introduced the concept of *localised* and *extended* states (fig. 3). Their characteristics can be summarised as follows:

- a) localised
 - i) confined to a finite region.
 - ii) normalisable.
 - iii) does not contribute to transport.
- b) extended
 - i) spread over the whole sample.
 - ii) un-normalisable.
 - iii) contributes to transport.

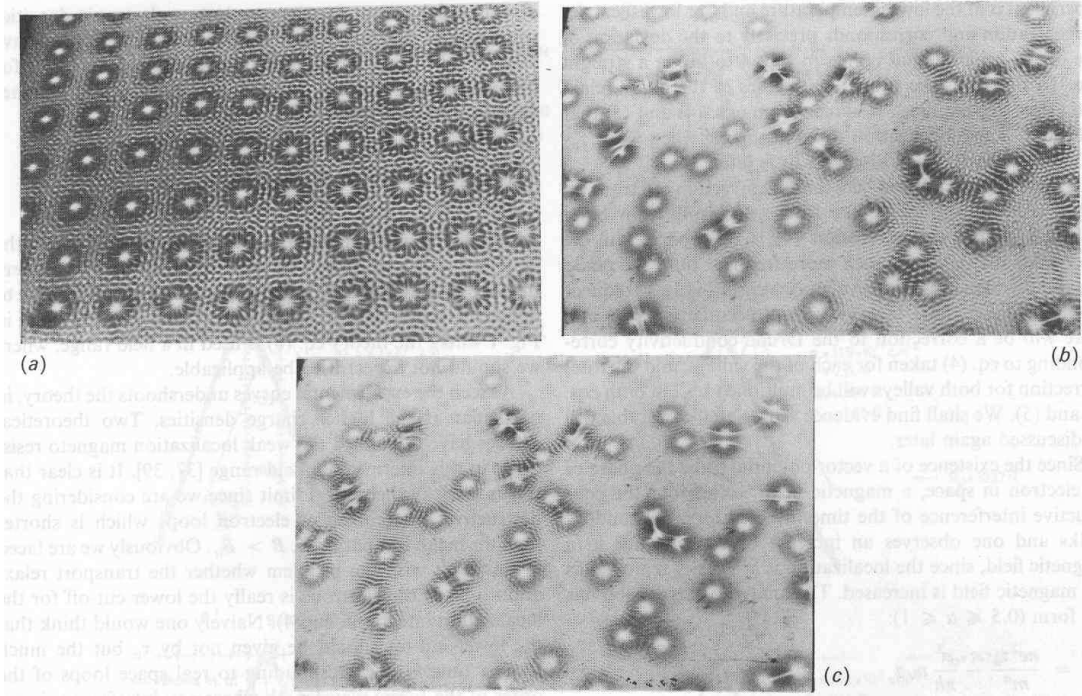


Figure 4: 3 Photos of a water bath exposed to an audio–frequency oscillation (not stroboscopic). (a) shows a situation where the obstacles sit in a regular quadratic lattice (frequency 76Hz). We see strong Bragg reflection corresponding to standing waves. (b) and (c) show randomly spaced obstacles exposed to two different audio frequencies (105Hz and 76Hz). Both (b) and (c) show standing wave patterns, but localised in different areas (with the permission of the authors from (Lindelof et al. 1986)).

It is worth noting at this point that the phenomenon of localisation is not confined to electrons, but can also be observed in other wave phenomena in random media, such as acoustic and optical waves, as well as (e.g.) water waves.

Later Mott (1968) (fig. 5) introduced the concept of a *mobility edge*. He argued that it is meaningless to consider localised and extended states which are degenerate since any linear combination of a localised and an extended state must be extended. Thus the concept of localisation can only be meaningful if there are separate energy regions of localised and extended states, rather like bands and gaps. These regions are separated by a *mobility edge* (fig. 6). Mott further argued that the states close to a band edge are more likely to be localised than those in the middle of a band. Since the localised states do not take part in conduction, electrons in a disordered semiconductor must be activated beyond the mobility edge rather than simply to the band edge. This should reveal itself in the slope of an Arrhenius plot,

$$\ln \sigma = \ln \sigma_0 - \frac{E_\mu - E_F}{k_B T} \quad (1)$$

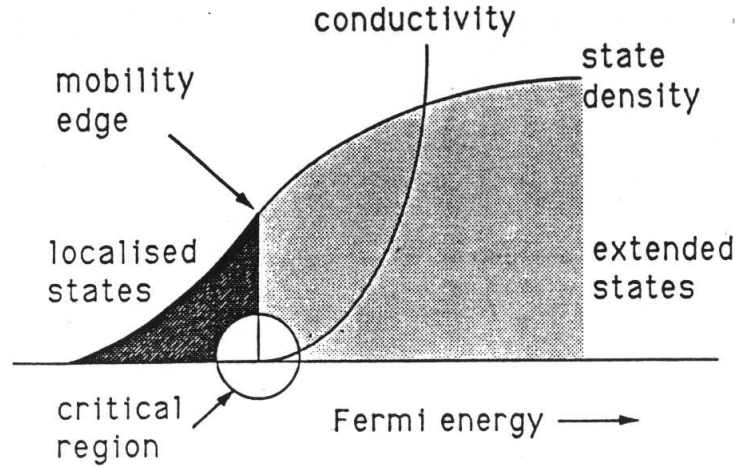


Figure 5: Nevill Mott (†1996)

Figure 6: Schematic illustration of the density of states and the conductivity near a band edge in a disordered system. The region of localised states is marked by the darker shading and no conductivity. The region of extended states is marked by lighter shading and a finite conductivity.

where E_μ and E_F are the mobility edge and Fermi energy respectively and σ is the conductivity.

2.3 Variable Range Hopping

At very low temperatures when activated conductivity is not significant and the Fermi level is in a region of localised states (e.g. an amorphous semiconductor) transport takes place by *hopping* between localised states. The electrons can gain or lose energy of order $k_B T$ by interaction with phonons and other excitations. In more than 1 dimension the electron is more likely to find a state in this energy range the further it hops.

On the other hand the exponential envelopes of the localised states must overlap for the phonon to couple them. Mott & Davis (1979) argued that the smallest energy difference in a range R is related to the density of states by

$$\Delta E = \left(\rho \frac{4}{3} \pi R^3 \right)^{-1}, \quad (2)$$

so that the probability of the electron hopping a distance R may be written

$$P(R) \propto \exp(-\alpha R - \Delta E/k_B T). \quad (3)$$

The distance of the most likely hop may be obtained by minimising the exponent with respect to R . In this way Mott & Davis (1979) found the famous $T^{-1/4}$ law. More precisely Mott's law is written

$$\sigma = \sigma_0 \exp \left(- \left[\frac{T_0}{T} \right]^{1/(d+1)} \right) \quad (4)$$

where d is the number of dimensions. This result has been verified many times in different systems. Or has it? In order to make an accurate measurement of the exponent, $1/(d+1)$, the conductivity must be measured over several decades of temperature while still remaining below the onset of activated transport. Thus the exponent cannot be determined very precisely. In

addition the measured value of the pre-exponential factor, σ_0 , often disagrees with the theory by several orders of magnitude.

Problem 2 Derive an expression for T_0 in Mott's law.

2.4 Minimum Metallic Conductivity

Yet another idea from Nevill Mott. The semi-classical conductivity can be written in the form

$$\sigma = \frac{ne^2\tau}{m} = \frac{ne^2\ell}{mv_F} = \frac{ne^2\ell}{\hbar k_F} \quad (5)$$

where n is the density of conduction electrons, m and e are the electron mass and charge respectively, τ is a scattering time, ℓ is the mean free path, and v_F and k_F are the Fermi velocity and wave vector. The density n of electrons is roughly proportional to k_F^d . The Ioffe & Regel (1960) criterion states that the wavelength cannot be greater than the mean free path ℓ (essentially the Heisenberg uncertainty principle) and, in any case, neither can be less than the inter-atomic distance, a . Hence the conductivity cannot be less than

$$\sigma_{\min} \propto \frac{e^2}{h} k_F^{d-1} \ell \geq \frac{e^2}{h} k_F^{d-2} \geq \frac{e^2}{h} a^{2-d} \quad (6)$$

In 2d k_F and a do not appear and σ_{\min} may be a universal constant. There have been many exper-



Figure 7: Mike Pepper

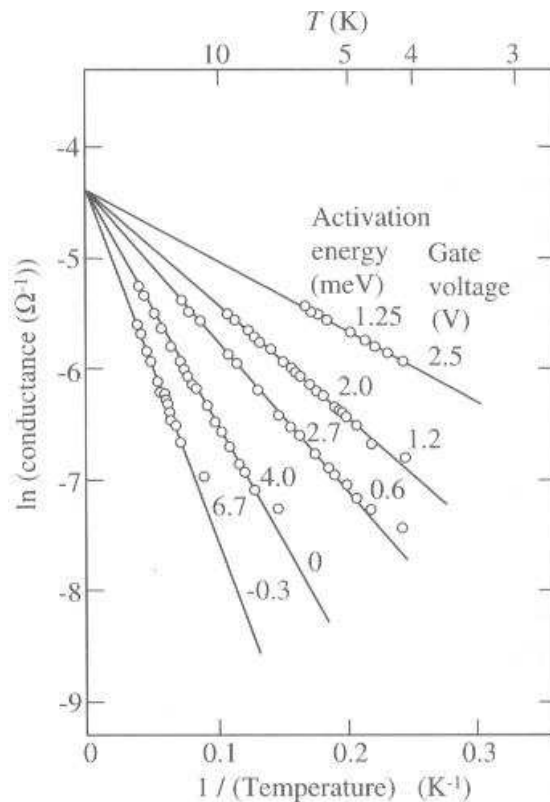


Figure 8: Arrhenius plot on a Silicon MOSFET, for various gate voltages V_g/V . Note the common intercept of the lines (Pepper 1978).

iments which purported to measure a σ_{\min} corresponding to $30\,000\,\Omega$ in 2d systems. By plotting (1) above for several different systems there appears to be a common value of $\ln \sigma_0$. This is now believed to be characteristic of an intermediate regime and not a proof of the existence of σ_{\min} (fig. 8).

2.5 The Coulomb or Fermi Glass

We now consider a system of localised states randomly distributed in energy but with no overlap between them. We do allow the electrons on these states to interact however. What happens when we move an electron from one state to another? In such a problem we have to be very careful about our definition of the energies of individual electrons.

- Unoccupied states: the energy required to add an electron to the state from outside the system.
- Occupied states: the energy lost by removing an electron from the state (or adding a hole).

Moving an electron is equivalent to removing the electron and bringing it back again, but we have to be careful that the electron is added to an $N - 1$ electron state rather than an N electron one. The net effect is that there is an extra contribution to the energy when an electron is moved. This extra contribution has the form of a Coulomb interaction between the electron in its new position and the hole left behind.

Consider now an electron moved from the highest occupied state to the lowest unoccupied state. The energy of the electron in its new state is the difference between the energy of the old state and that of the new state (i.e. almost zero) plus the contribution from the electron–hole interaction. As the extra term is negative it would appear to be possible for the new state to have lower energy than the old one. As the system was supposed to start in its ground state there cannot be such a new state. Efros & Shklovskii (1975) argued that there must therefore be a

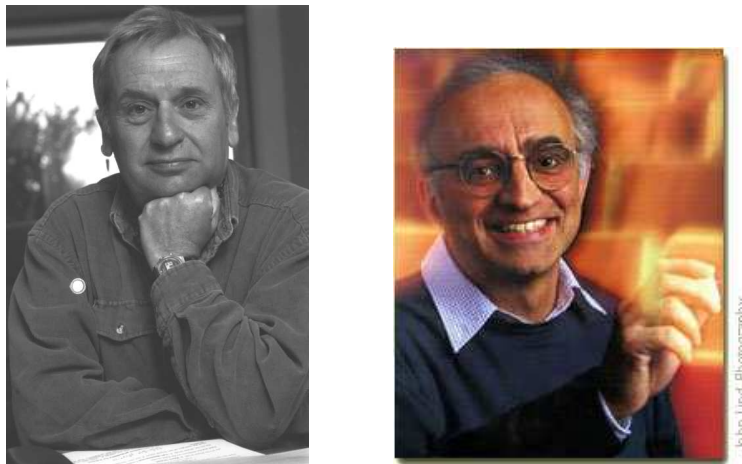


Figure 9: Alexei Efros (left) and Boris Shklovskii (right)

soft gap in the density of states around the Fermi level. Consider a small interval ϵ around the Fermi level. Typically we have to look in a volume

$$R^d = \frac{1}{\epsilon \rho(E_F)} \quad (7)$$

to find 2 states separated by ϵ . On the other hand such states must be spatially separated by at least

$$R = \left(\frac{e^2}{4\pi\epsilon_0} \right) \frac{1}{\epsilon} \quad (8)$$

We can eliminate R from (7) & (8)

$$\epsilon^d \propto \epsilon \rho(E_F \pm \epsilon) \quad (9)$$

which clearly only makes sense if the density of states, ρ , takes the form

$$\rho(E) \propto |E - E_F|^{d-1} \quad . \quad (10)$$

Thus the DoS touches zero at $E = E_F$, a *soft gap*, sometimes called the Coulomb gap or Efros & Shklovskii gap.

Problem 3 *The above analysis assumes that we are using the 3 dimensional form for the Coulomb interaction, the solution of Poisson's equation in 3D. Show that if we consider the 1D case with the 1D Poisson's equation we obtain a singularity rather than a soft gap.*

Problem 4 *The analysis of section 2.3 assumes that the density of states at E_F is constant. Repeat that analysis for $\rho \propto \epsilon^\delta$ and obtain an expression for the temperature dependence of the conductivity. Hence show that when the DoS takes the Efros & Shklovskii form the temperature dependence of the conductivity is independent of dimension d .*

3 Scaling Theory

3.1 The Thouless Energy

Suppose we consider our disordered system as a single unit cell of a much larger crystalline system. Alternatively we apply periodic boundary conditions to our system. The group velocity of an electron is given by

$$v_g = \hbar^{-1} \frac{\partial E}{\partial k} \quad (11)$$

The energy of a state will be sensitive to \mathbf{k} only if it's amplitude is finite from one side of the system to the other, or all the way round a ring. Otherwise the change of phase can be gauge transformed away, i.e. by applying a transformation of the sort $\psi(\mathbf{r}) \mapsto \psi(\mathbf{r}) \exp(i\theta(\mathbf{r}))$. This can easily be understood by considering the system with periodic boundary conditions matched on to a torus. In this case the function $\theta(\mathbf{r})$ must match up when it is followed once round the system. In a localised system the amplitude falls to zero and the phase becomes irrelevant, so that the change of \mathbf{k} can always be transformed away. In this case the energy, E , cannot depend on \mathbf{k} and the group velocity must be zero.



Figure 10: David Thouless

David Thouless (1974) developed this sort of argument in a slightly different way. The time taken for an electron to diffuse across a (hyper)cube is given by

$$\tau = L^2/D \quad (12)$$

where L is the length of an edge of the system and D is the diffusion constant. As τ may also be interpreted as the time spent by the electron in the cube there is an associated energy uncertainty

$$\delta E = \hbar/\tau \quad . \quad (13)$$

We now define ΔE to be the average spacing between levels, which is related to the density of states, $\rho(E)$, by

$$\Delta E = \left(L^d \rho(E) \right)^{-1} \quad , \quad (14)$$

as we have used twice already. We can now consider what happens when similar but different cubes are joined together. This gives us 2 regimes:

$\delta E < \Delta E$ There is no overlap between levels in neighbouring cubes; the electron cannot travel from one cube to another; the states are localised.

$\delta E > \Delta E$ Levels from different cubes overlap; the electron finds it easy to travel from one to another; the states are extended.

In order to make a connection between the diffusion and the conductivity we invoke the Einstein relation

$$\sigma = e^2 D \rho(E) \quad . \quad (15)$$

By combining these equations we obtain

$$\delta E = \frac{\hbar}{e^2} \frac{\sigma}{\rho L^2} = \frac{\hbar}{e^2} \sigma L^{d-2} \Delta E = g \Delta E \quad , \quad (16)$$

where

$$g = \frac{\hbar}{e^2} \sigma L^{d-2} \quad (17)$$

is the dimensionless conductance (inverse resistance).

The change in behaviour when 2^d systems are joined together can be described solely in terms of $g = \delta E/\Delta E$, the Thouless number (δE is sometimes called the *Thouless energy*).

$$g(2L) = f(g(L)) \quad (18)$$

Related to this is the fact that the spectrum of extended states is said to be *continuous* whereas that of localised states is *discrete*. In a continuous spectrum a finite fraction of a range of energies constitutes eigenenergies in an infinite system. In a discrete spectrum, by contrast, only discrete energies are eigenenergies and the fraction of a range of energies which are eigenenergies goes to zero for an infinite system, even if the number of such energies is itself infinite. Compare this with the set of rational numbers.

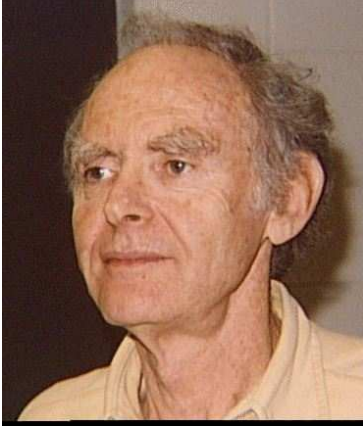


Figure 11: Elihu Abrahams

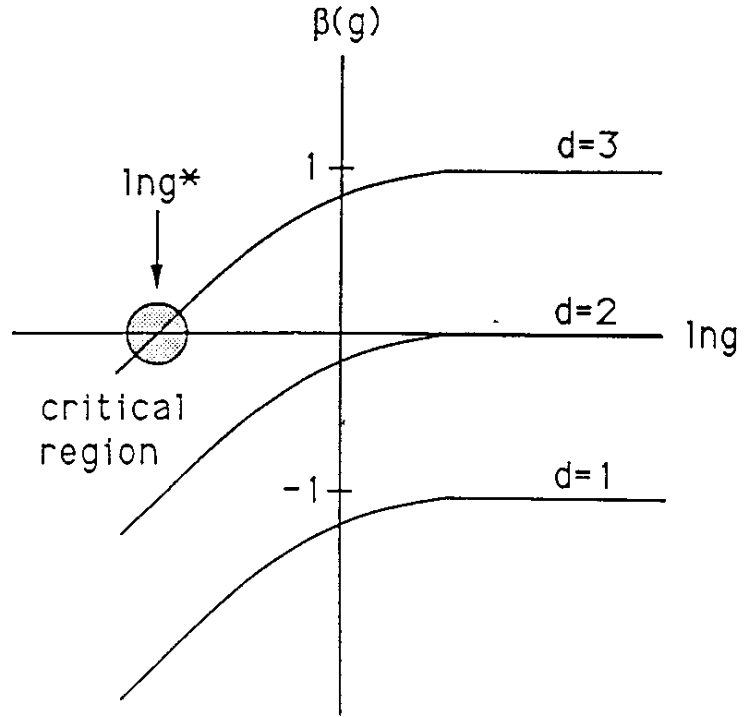


Figure 12: β function for the conductance g in 1, 2 and 3 dimensions.

3.2 The Gang of Four

A decisive breakthrough in the theory of the metal insulator transition was made in 1979 when Abrahams et al. (1979) published their scaling theory, who picked up Thouless' idea and recast it as a differential equation

$$\frac{d \ln g}{d \ln L} = \beta(\ln g) \quad . \quad (19)$$

Note that the differentiation is with respect to $\ln L$ rather than L as the basic transformation being described involves *multiplication* of L by a constant (2 in the simple derivation) rather than *addition* of a small constant.

What then are the properties of the function $\beta(\ln g)$? For strong disorder, the states are highly localised and we expect g to fall exponentially with the size of the system. Thus

$$g = g_0 \exp(-\alpha L) \quad (20a)$$

$$\beta(\ln g) = \ln g - \ln g_0 \quad (20b)$$

In the case of weak disorder the classical behaviour should be valid

$$g = \frac{\hbar}{e^2} \sigma L^{d-2} \quad (21a)$$

$$\beta(\ln g) = d - 2 \quad (21b)$$

where σ is the conductivity of the material. Thus, for $g \ll 1$, $\beta(\ln g)$ is always negative. For $g \gg 1$, $\beta(\ln g)$ is positive in 3 dimensions, negative in 1-d and zero in 2-d.

Since g and β should be smooth functions of disorder, energy etc. β must change sign in 3-d, may change sign in 2-d and probably does not in 1-d (fig. 12).

What then is the meaning of the sign of β ? If β is negative g is decreasing with increasing L , β becomes even more negative, and so on until g is zero. That is: there is no conductance. When β is positive, by contrast, g is increasing with increasing L and will eventually approach the classical behaviour, (21). Thus there is no metal–insulator transition in 1–d. All states are localised, just as in the percolation case.

3.2.1 Three Dimensions

In 3D there is no doubt that there is indeed a transition. In order to understand the nature of the transition let us examine the β –function more carefully, especially close to the point $\beta = 0$. Here we may approximate (19) as

$$\frac{d \ln g}{d \ln L} = \beta' (\ln g - \ln g^*) \quad \text{where} \quad \beta' = \left. \frac{d \beta(\ln g)}{d \ln g} \right|_{g=g^*} . \quad (22)$$

This differential equation has a solution

$$\ln g = \ln g^* + AL^{\beta'} \quad (23)$$

where A is a constant of integration and must be linear in the energy or disorder, i.e. $A \sim (E - E_c)$ so the $g = g^*$ when $E = E_c$.

On the other hand we can look for a formal solution of (19) without an approximation by considering

$$\frac{d \ln g}{\beta(\ln g)} = d \ln L \quad (24a)$$

$$F(\ln g) = \ln L - \ln L_0 \quad (24b)$$

$$\ln g = f(L/L_0) \quad (24c)$$

here $f(x)$ is a function which may be found in principle by integrating the left-hand-side of (24a) and functionally inverting the result. By comparing (23) and (24c) we can deduce that the length scale L_0 must depend on the energy or disorder as

$$L_0 \propto |E - E_c|^{-1/\beta'} . \quad (25)$$

On the localised side of the transition we identify L_0 with the localisation length. What does it mean on the extended side? We note that for weak disorder we compare (24c) with (21a) and deduce that the conductivity $\sigma \sim L_0^{-1}$ in 3D.

3.2.2 Two Dimensions

What about 2–d? This is the marginal case. The existence of a transition depends on whether $\beta(\ln g \rightarrow \infty)$ approaches zero from above or from below. Abrahams et al. (1979) were able to show that the leading term in an expansion of (19) in $1/g$ has the form

$$\beta(\ln g) = -\frac{a}{g} \quad (26)$$

Thus β is always negative and all states are localised. There is no true metallic conductivity in 2 dimensions.

While it is true in a strict mathematical sense that everything is localised in 2–d, this result needs interpretation. After all there is no shortage of experimental evidence that there is considerable conductivity in some 2–d systems (eg HEMT's).

There have been very few attempts to calculate the actual numerical value of the localisation length for a real 2-d system. Results of computer simulations suggest that it can be of order cm even when the potential fluctuations are of the same order as the band width (MacKinnon & Kramer 1983a). If we ask for the localisation length in a very pure sample, numbers larger than the universe tend to emerge, (eg $10^{10^{30}}$)(MacKinnon & Kramer 1983b). Is it meaningful then to talk about localisation in this case, when the localisation length is often much larger than the sample size?

Problem 5 *A certain system is described by a β -function of the form $\beta(g) = 1 - a/g^n$. What is the value of the corresponding critical exponent?*

3.3 Experiments on Weak Localisation

In practise, although the localisation in 2D often cannot be observed directly there are various precursor effects which can be observed fairly easily. These are collectively referred to as *weak localisation*. By integrating (26) above we obtain a formula for the conductivity

$$\sigma = \sigma_0 - a \ln L \quad (27)$$

This is still not of much use to us. The sample size is not an easily varied quantity. However the discussion so far has been only in terms of disorder effects or elastic scattering. Inelastic effects such as scattering by phonons and by other electrons must also be considered. There is one important distinction between elastic and inelastic scattering. In elastic scattering there is a well defined phase relationship between an incident and a scattered wave, whereas inelastic scattering destroys such phase coherence. Since the localisation is really an interference phenomenon, it can be destroyed by inelastic scattering. This can be built into our picture in a simple way. Equation (27) is valid until the electron is scattered in-elastically, so we can identify the length L with the inelastic scattering length L_{inel} . In general $L_{inel} \propto T^{-\alpha}$, so that we can substitute in (27) to obtain

$$\sigma = \sigma_0 + \alpha a \ln T \quad (28)$$

This logarithmic temperature dependence has been observed in a number of systems, MOS-FET's, thin films etc. and was considered a confirmation of the concept of weak localisation in 2-d.

4 Quantum Interference

What is the origin of the negative coefficient in (26)? In order to understand this we first consider a simpler problem: that of quantum interference around a ring. Our bulk system will then be treated as an ensemble of such rings.

Consider then a disordered ring with contacts at two points diagonally opposite each other. We assume that the ring is large compared to the mean free path but small compared to the localisation length. There is lots of scattering but no localisation.

An electron which starts at one contact can travel to the other contact by one of two routes (fig. 13a). The two waves which arrive at the other side have been scattered differently. There is therefore no particular phase relationship between them. On the other hand, if we follow the two waves right round the ring and consider the effect back at the origin, then the two waves have been scattered identically (fig. 13b). One is the time reversed case of the other. They thus arrive back at the origin with the same phase. The probability of returning to the origin is twice what it would be if we had ignored the interference and simply added the intensities.

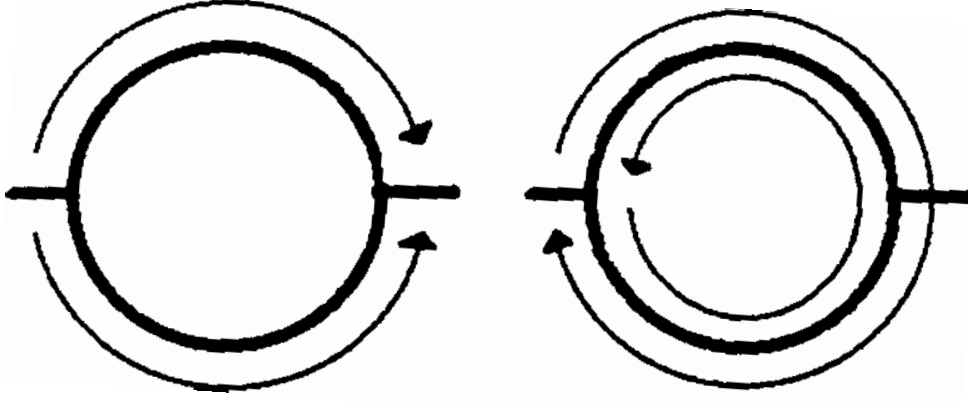


Figure 13: Scattering paths round a ring. (a) waves round opposite sides of the ring interfere randomly on the other side as the paths travelled are different. (b) waves round the whole ring but in opposite directions interfere constructively back at the origin as the paths travelled are the same.

If we consider the probability that a particle in a general disordered system will eventually return to the origin, the formula will contain terms which refer to pairs of waves which go round rings in opposite directions, in addition to lots of other terms. These other terms, at least for weak scattering, don't do anything non-classical, but the ring interference terms still give an enhanced probability of returning to the origin, and thus a tendency toward localisation.

The same interference phenomenon can also be described in k -space. In this case a wave which starts with wave vector \mathbf{k} has a higher probability of being scattered to $-\mathbf{k}$ than to any other direction (fig. 14). This *correlated back scattering* has given rise to a number of experiments looking for optical analogies of weak localisation (fig. 15).

To avoid any misunderstanding it should perhaps be pointed out that backscattering is different from specular reflection. In the latter case only the component of \mathbf{k} perpendicular to the surface is reversed, whereas in the weak localisation effect all the components are reversed.

4.1 Negative Magneto-resistance, Aharonov–Bohm & Sharvin–Sharvin

What happens when we introduce a magnetic field into this system? Consider again a single ring. When a magnetic flux, $\Phi = \pi R^2 B$, is fed through a ring of radius R the Schrödinger equation is changed by

$$\mathbf{p}^2 \mapsto (\mathbf{p} - e\mathbf{A})^2 \quad (29)$$

where \mathbf{A} is the magnetic vector potential. Choosing polar coordinates and an appropriate gauge, $\mathbf{A} = \frac{1}{2}Br\hat{\theta}$, (29) can be rewritten in the form

$$\mathbf{p}^2 \mapsto \left(\frac{\hbar}{i} \frac{1}{R} \frac{\partial}{\partial \theta} - \frac{1}{2}eBR \right)^2 \quad (30)$$

which still has eigenfunctions of the form $\exp(im\theta)$ but $\pm m$ are no longer degenerate. Thus the two opposite paths round the ring are no longer equivalent, the probability of a particle returning to the origin is no longer enhanced and there is no enhanced back scattering. Translated into the language of a solid rather than a single ring, we see that the chief mechanism which leads to

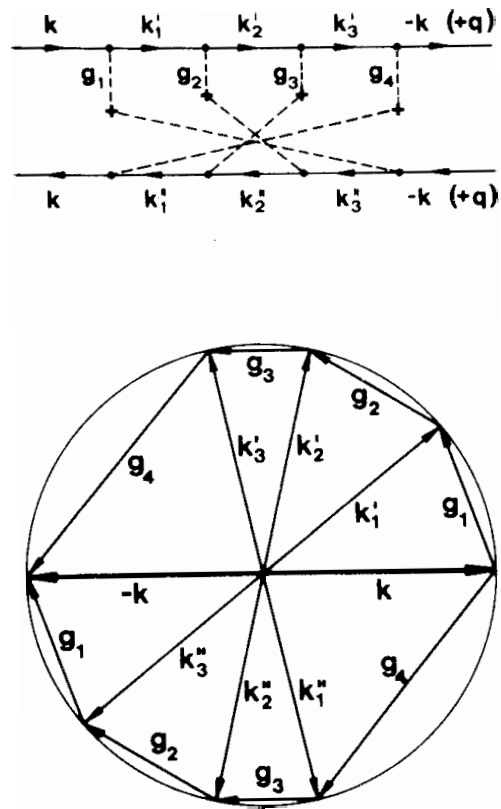
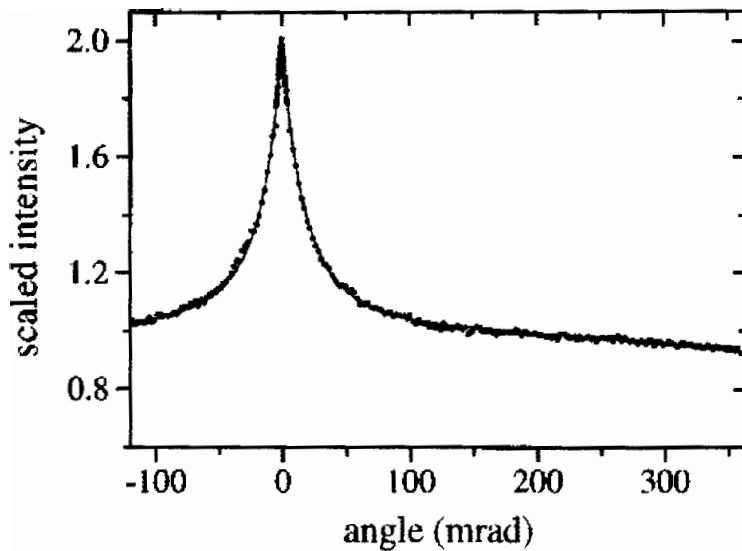


Figure 14: The totally crossed Feynman diagram (upper picture) and its physical interpretation (lower picture). Among the many scattering path around the Fermi surface from there are pairs consisting of the same steps but in reverse order. These interfere constructively at and this is observed as enhanced back scattering.



<http://www.wins.uva.nl/research/scm/adlag/articles/aftlife.htm>

Figure 15: Intensity of scattered light from a white-paint suspension as a function of angle around the backscatter direction (from (Lagendijk 1994)). The enhancement at exactly the backscattering direction (0 mrad corresponds to 180 degrees) is clearly visible).

weak localisation is no longer active. Hence the resistance decreases in a magnetic field; hence negative magneto-resistance (see fig. 19).

Returning though to the single ring. It so happens that it is now possible to make small rings or cylinders with dimensions such that these phenomena can be observed directly.

Firstly, consider again the interference between two waves which cross to the opposite side of the ring by different routes. The phase difference between the two waves will be

$$\theta_0 + 2\pi R \frac{1}{2} eBR/\hbar = \theta_0 + \frac{e}{\hbar} \Phi \quad (31)$$

This has the value $\theta = \theta_0 + 2n\pi$ whenever $\Phi = n \frac{\hbar}{e}$. Therefore we expect the current through the ring to vary with a period of one flux quantum¹. This is one variant of the well known Aharonov-Bohm (Aharonov & Bohm 1959) effect. In fact when Sharvin & Sharvin (1981) performed the experiment on a hollow magnesium cylinder the period was found to be $2\hbar/e$ (fig. 16). Why? It certainly does not indicate pairing of electrons.

In fact Altshuler et al. (1981) showed that there is an alternative interference process with precisely this period. As before we must consider interference back at the origin. In this case the total flux enclosed is doubled. However, since the two paths are identical in the absence of a magnetic field the term θ_0 vanishes. By contrast in the Aharonov-Bohm effect θ_0 is non zero with a random value from sample to sample. Since Sharvin & Sharvin's cylinder may be considered as an ensemble of such rings, the phase is randomised between samples and no oscillation of the resistance is observed.

¹Here the flux quantum is \hbar/e rather than $\hbar/2e$ as we are dealing with a single electron effect rather than one due to pairs of electrons, such as superconductivity

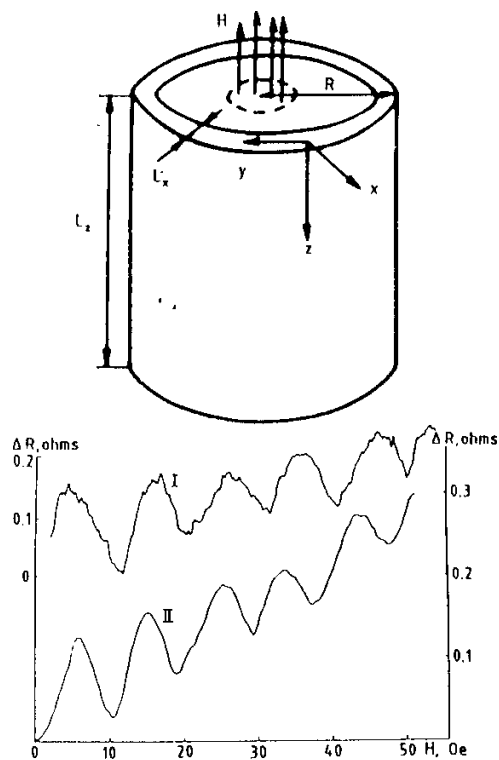


Figure 16: Aharonov–Bohm like magneto-conductance oscillations observed in normally conducting Mg cylinders of diameter $1.5 \mu\text{m}$ by Sharvin & Sharvin (1981). Left and right resistance scales correspond to samples 1 and 2, respectively. The periodicity of the oscillations corresponds to $\Delta\Phi = h/2e$.

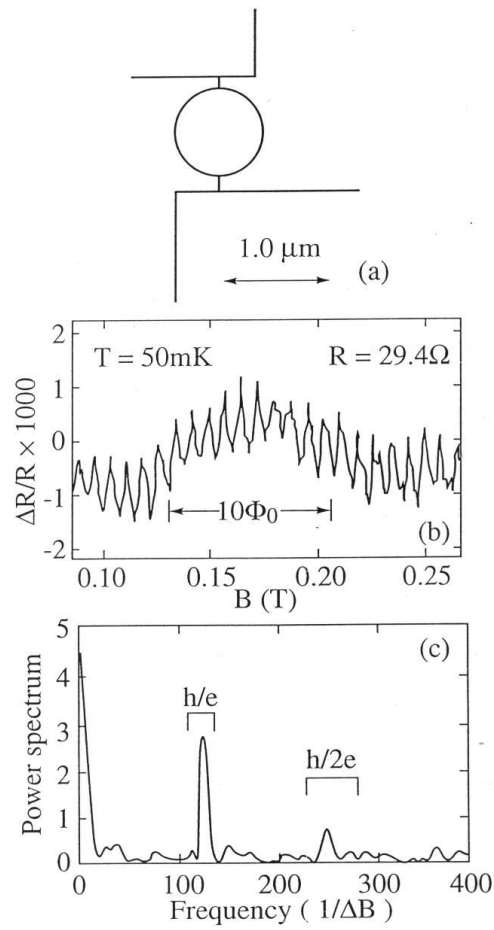


Figure 17: Aharonov–Bohm Oscillations in a small ring. The period of the oscillations is 1 flux quantum (i.e. $\Delta\Phi = h/e$) (from (Umbach et al. 1987)).

Problem 6 Consider a single isolated ring made of 1-dimensional material and threaded by a magnetic flux Φ . Imagine that we can unroll the ring and lay identical copies of it end to end to form an infinite crystal. The eigenstates of the crystal obey Bloch's theorem with a wavevector k . Show that the eigenenergies of the ring as a function of Φ are identical to those of the crystal as a function of k . Hence show that in a finite flux the ring may carry a current. How might such a current be measured? This phenomenon is known as a "persistent current".

4.2 Single Rings, Non-Local Transport

More recently it has become possible to etch very fine patterns on metal films leaving very fine wires. An example is shown in fig. 17. Note the periodic oscillations of the current as a function of the magnetic fields with the period of a single flux quantum in contrast with the Sharvin & Sharvin experiment.

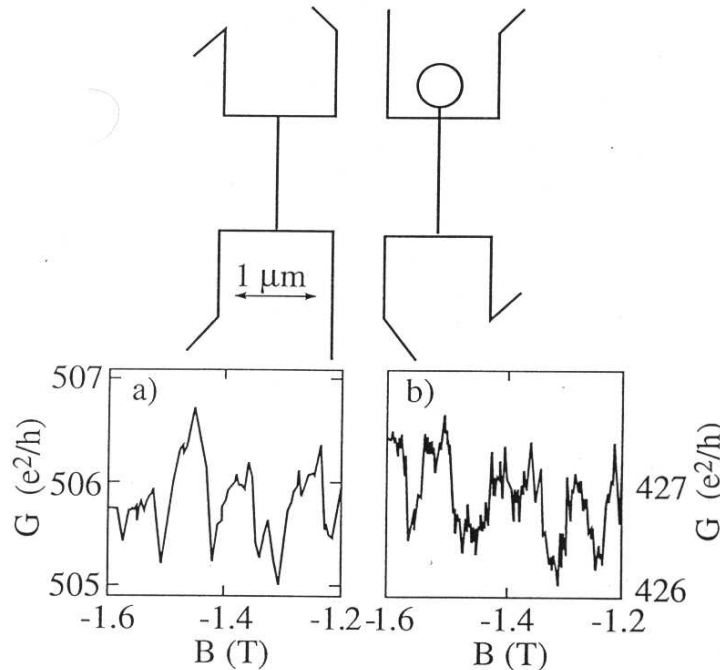


Figure 18: Non-local transport in thin wires. In (a) only random fluctuations are observed. In (b) however interference round the "head" contributes a periodic oscillation (from (Umbach et al. 1987)).

An even more dramatic example of quantum interference effects on transport in microstructures can be seen in fig. 18. Here the figure without the ring (or head) shows universal conductance fluctuations (see below) whereas in the second figure, with a ring, a periodic oscillation is superimposed. Note that classically the ring is irrelevant as it constitutes a dead end for the current but that quantum interference between different paths round the ring can still contribute, leading to the oscillations. This is the first of a range of phenomena involving *non-local* transport, in which classically irrelevant interference paths can contribute to transport.

Consider a sample with several different leads in which a current is sent between leads m and n and a potential difference is measured between leads i and j . The result may be defined in terms of a generalised resistance $\mathcal{R}_{ij,mn}$ such that

$$V_{ij} = \mathcal{R}_{ij,mn} I_{mn} \quad (32)$$

This is a very general notation for describing most common transport measurements, and is often used to represent non-local effects.

4.3 Spin-Orbit Coupling, Magnetic Impurities, etc.

Soon after the discovery of weak localisation the quantum Hall effect (see later) was discovered. It rapidly became clear that (26) is not universally valid. In fact there are three important exceptions. Besides high magnetic fields (see below) spin-orbit coupling and magnetic impurities will also give deviations.

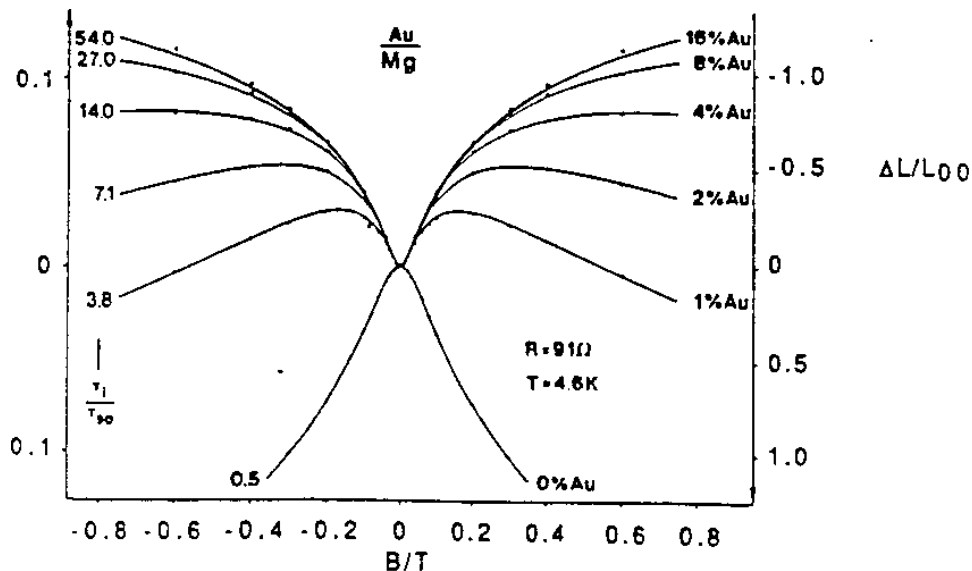


Figure 19: The magneto-resistance ΔR of thin Mg-films. The clean film shows a negative magneto-resistance indicating localisation. When the film is covered with a small amount of gold atoms the magneto-resistance becomes positive due to increasing spin-orbit scattering. The right scale shows the magneto-conductance ΔL . On the left, the ratio of the inelastic scattering time and the spin-orbit scattering time is indicated (redrawn after Bergmann (1984)).

In the case of spin-orbit coupling the deviation from the classical behaviour is positive. In fact, it has not been shown analytically that there can be any localised states for purely spin-orbit scattering. In a beautiful set of experiments on Au doped Mg Bergmann (1984) was able to demonstrate the validity of the perturbation theory for such systems (fig. 19). (n.b. spin-orbit scattering rises as Z^4). The spin-orbit effect is sometimes termed *weak anti-localisation*.

Magnetic impurities, destroy the weak localisation effect by destroying the time reversal symmetry. The β -function has a leading term $-a/g^2$, and the localisation is even weaker than before. Magnetic impurities can also give rise to a divergence of resistance at low temperatures known as the Kondo (1969) effect (Hewson 1993).

4.4 Universal Conductance Fluctuations

So far in our discussion of the conductivity of disordered systems, we have implicitly assumed that the transport properties are *self-averaging*. As is usual in statistical mechanics, we have assumed that the sample is so big that the distribution of possible values of the resistance is very narrow, essentially a δ -function. In fact this assumption is invalid. As long as we are working in a regime where the inelastic scattering length is larger than the sample size, the

sample cannot be considered as made up of a large number of statistically independent systems. A small change in one place may have consequences for the whole sample.

In the mesoscopic regime defined earlier the conductance is a number of order e^2/h . In fact at $T \rightarrow 0$ the standard deviation $(\delta g)^2$ is also of order e^2/h . Note that this behaviour is exactly what would be expected if the conductance takes the value e^2/h or zero randomly.

Experimentally this can be measured, not by comparing different samples, but by looking at the way the conductance depends on such quantities as gate voltage or magnetic field. In the former case the Fermi level is moved through the spectrum, alternately seeing states and gaps. In the latter the spectrum is moved in a systematic way.

The results look like noise (fig. 18a). Unlike noise however the structure is reproducible. Different samples behave qualitatively similarly but differ in details.

The apparently random fluctuations also have a characteristic width which is related to the Thouless energy.

4.5 Ballistic Transport

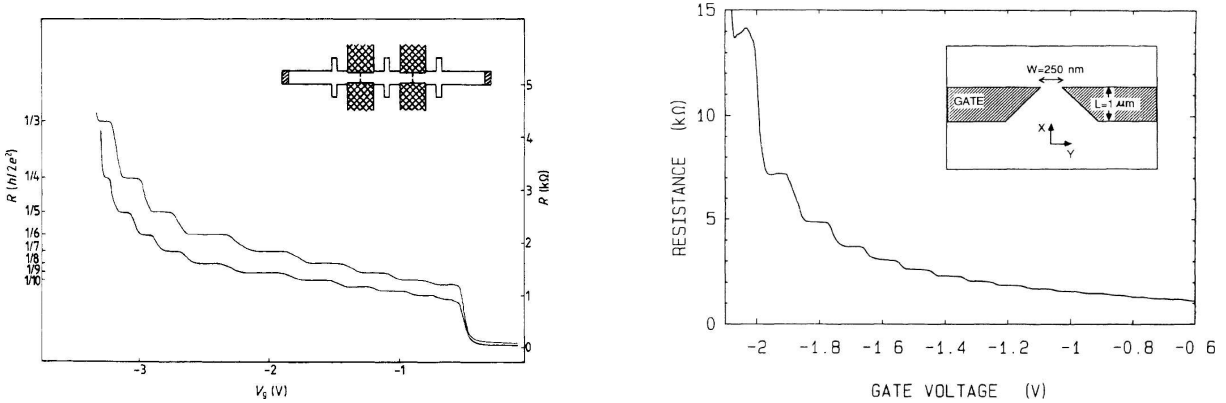


Figure 20: a) Resistance of a single-nanoconstriction device at $T = 0.1$ K plotted against voltage V_g . The two curves have different carrier concentrations induced by sample illumination. Note the quantised resistance values in units of $h/2e^2$. The inset shows a diagram of the device with 2 split gates, one of which is used in the experiment. The split gate is $0.5\mu\text{m}$ wide and $0.4\mu\text{m}$ long (from (Wharam et al. 1988)). b) Conductance of a nanoconstriction created by a split gate against gate voltage, at $T = 0.6$ K. The conductance (after subtraction of the lead resistance) shows clear plateaux quantised at multiples of $2e^2/h$. Inset: diagram of the device with the split gate shown as the dashed region (from (van Wees et al. 1988)).

By using a so-called *split gate* it is possible to study the transition from 2-dimensional to 1-dimensional behaviour. If the scattering is weak and the 1-d channel short enough it is possible to measure a remarkable new phenomenon. When the resistance of such a channel is measured as a function of electron density (i.e. conventional gate voltage) it is found to be quantised as (fig. 20)

$$R = \frac{h}{2e^2n} \quad G = \frac{2e^2}{h}n \quad (33)$$

To understand why this occurs consider a simple one dimensional system. We wish to measure the current flowing along the 1-d chain as the voltage at one end is raised by a small amount δV . The current flowing is given by the product of the number of states in the interval δV and

the current carried per state. That is

$$I = \left(\frac{e\delta V}{2\pi(\partial E/\partial k)} \right) \left(e \frac{1}{\hbar} \frac{\partial E}{\partial k} \right) \quad (34a)$$

$$= \frac{e^2}{h} \delta V \quad (34b)$$

Note that the terms, $\partial E/\partial k$, which depend on the detailed electronic structure, cancel leaving a result that does not depend on E or on k . In a quasi-1-d system of finite width there will be a contribution to the conductance from each occupied transverse state of the channel leading to (33) where we can now interpret the integer n as the number of occupied channels, and the extra factor 2 is due to spin degeneracy.

Another unexpected result is found when 2 such *ballistic channels* are measured in series. The resulting conductance is simply the lower of the 2 conductances. The resistances do not add.



Figure 21: Rolf Landauer (†1999) and Markus Büttiker.

5 The Büttiker or Landauer–Büttiker formula

The simple concepts given above can be generalised using an approach originally due to Landauer (1970) and generalised by Büttiker (Büttiker et al. 1985, Büttiker 1986) and generalised to deal with superconducting contributions by Lambert (1991).

5.1 The Landauer formula

We start by considering a simple case in which the barrier disordered system is embedded between perfect 1-dimensional leads which are in turn embedded between reservoirs (fig. 22). The left and right reservoirs have chemical potentials μ_1 and μ_2 and we assume $\mu_1 > \mu_2$ so that we expect the current to flow from left to right. We also assume that the transmission and reflection coefficients of the barrier are independent of energy in the range $\mu_1 \rightarrow \mu_2$; in effect we are assuming that $\mu_1 - \mu_2 < \delta E$, the Thouless energy (section 3.1). The current emitted by the left reservoir in the range $\mu_1 \rightarrow \mu_2$ is given as by

$$I = ev_g \rho (\mu_1 - \mu_2) = 2 (e/h) (\mu_1 - \mu_2) \quad (35)$$

where the group velocity, v_g , cancels with the density of states, ρ . If only a fraction T of this is transmitted through the barrier we can write the net current through the whole system as

$$I = 2 (e/h) T (\mu_1 - \mu_2) \quad . \quad (36)$$

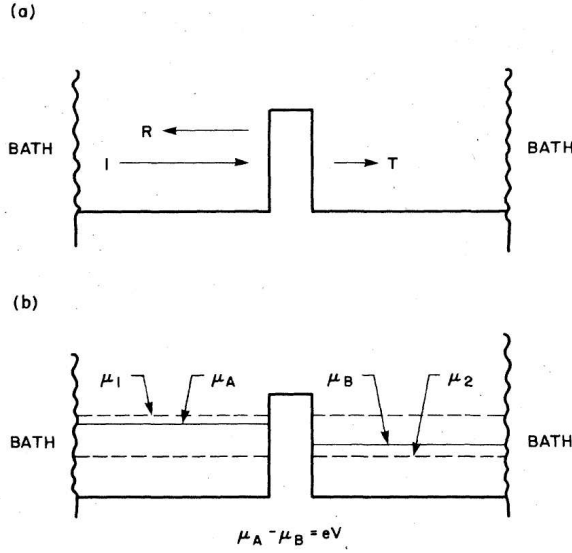


Figure 22: The geometry of (Landauer 1970). The obstacle is connected to two incoherent reservoirs by ideal 1D conductors. A stream of particles with unit density hits the barrier from the left, a fraction R is reflected, and a fraction T transmitted. (b) The chemical potentials in the single channel case. The LHS reservoir emits electrons up to a quasi-Fermi -energy μ_1 , and the RHS reservoir emits electrons up to a quasi-Fermi-energy μ_2 . μ_A and μ_B are the chemical potentials in the perfectly conducting leads to the left and the right of the barrier.).

It would appear that we already have our formula for the conductance. However, there is a problem: as $T \leq 1$ the conductance would appear to have a maximum value. Surely it should become infinite in the absence of scattering. In fact we have failed to take account of the contribution to the potential from the electrons which are transmitted through the system or those reflected from it. We can represent this effect by introducing chemical potentials for the leads, μ_A and μ_B which differ from those of the reservoirs. This is allowed as we may assume that the electrons are thermalised while passing along the leads. We have to be careful about the charge neutrality of the leads: this is guaranteed as long as the number of occupied states above μ_A (μ_B) is equal to the number of empty states below. The total number of states in the range $\mu_1 \rightarrow \mu_2$ is $2\rho(\mu_1 - \mu_2)$ (the 2 arises because there is a left-going and a right-going state at each energy). Consider now the right hand wire: the number of occupied states above μ_B is $T\rho(\mu_1 - \mu_B)$ and the number of unoccupied states below μ_B is $(2 - T)\rho(\mu_B - \mu_2)$. Thus μ_B is determined by the condition

$$T\rho(\mu_1 - \mu_B) = (2 - T)\rho(\mu_B - \mu_2) \quad , \quad (37)$$

which may be rewritten in terms of the reflection matrix R as

$$(1 - R)\rho(\mu_1 - \mu_B) = (1 + R)\rho(\mu_B - \mu_2) \quad . \quad (38)$$

We can use similar arguments on the left-hand lead to obtain

$$(1 + R)\rho(\mu_1 - \mu_A) = (1 - R)\rho(\mu_A - \mu_2) \quad . \quad (39)$$

By solving (38) and (39) we can express $\mu_1 - \mu_2$ in terms of $\mu_A - \mu_B$ as

$$\mu_1 - \mu_2 = (\mu_A - \mu_B)/R \quad . \quad (40)$$

In order to guarantee the charge neutrality of the leads the Fermi energies μ_A and μ_B must be the same distance above the bottom of the band and their difference must be accounted for by the electrical potential²

$$\mu_A - \mu_B = eV \quad . \quad (41)$$

Substituting (40) and (41) into (36) gives us an expression for the conductance

$$G = 2 \frac{e^2 T}{h R} \quad . \quad (42)$$

This is the original Landauer (1970) formula. Note that it becomes infinite in the absence of scattering as expected.

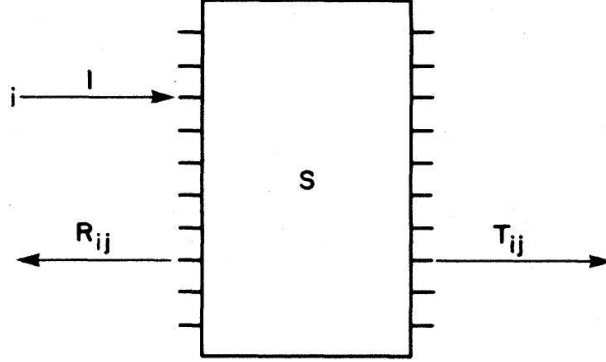


Figure 23: A multichannel system S . A unit current in channel i is reflected into channel j with probability R_{ij} and transmitted into j with probability T_{ij} . Indices i and j run from 1 to N . Input channels are incoherent with each other.)

5.2 Generalisation to finite cross section

There is a large and controversial literature on the generalisation of (42) to leads with a finite number of channels. Here we outline the approach due to Büttiker et al. (1985). In a lead with a finite cross-section the wave function is quantised in the transverse direction and the different *channels* may be labelled by the corresponding quantum numbers. Note, however, that the following derivation only requires us to consider *open* channels, i.e. those for which the energy associated with the transverse part is below the Fermi energy and which therefore have longitudinal parts in the form of plane waves rather than exponentials³. The crucial assumption is that the various μ 's are common to all the channels on one side of the barrier. We define T_{ij} and R_{ij} as in figure 23. The current transmitted into channel i on the right-hand-side is given by

$$I_i = 2 \left(\frac{e}{h} \right) \left[\sum_j T_{ij} \right] (\mu_1 - \mu_2) \quad . \quad (43)$$

which leads us to define

$$T_i = \sum_j T_{ij} \quad R_i = \sum_j R_{ij} \quad . \quad (44)$$

²strictly speaking μ_A and μ_B are *electrochemical* potential energies. We insist that the 2 *chemical* potentials must be equal and that the difference must therefore be a purely *electrical* potential

³Note that when calculating the transmission coefficients themselves it may not be valid to ignore evanescent channels. It is also important to distinguish between wave function amplitude and current transmission matrices.

Thus we can write an expression for the total current in the form

$$I_{\text{tot}} = \sum_i I_i = (\mu_1 - \mu_2) 2 \frac{e}{h} \sum_i T_i = (\mu_1 - \mu_2) 2 \frac{e}{h} \text{Tr } tt^\dagger \quad (45a)$$

$$= (\mu_1 - \mu_2) 2 \frac{e}{h} \sum_i (1 - R_i) \quad . \quad (45b)$$

We now define μ_A and μ_B as before and look for the number of occupied (unoccupied) states above (below) μ_A and μ_B to obtain generalisations of (37) and (39)

$$\sum_i T_i \rho_i (\mu_1 - \mu_B) = \sum_i (2 - T_i) \rho_i (\mu_B - \mu_2) \quad (46a)$$

$$\sum_i (1 + R_i) \rho_i (\mu_1 - \mu_A) = \sum_i (1 - R_i) \rho_i (\mu_A - \mu_2) \quad , \quad (46b)$$

which leads directly to

$$G = 2 \frac{e^2}{h} \frac{2 (\sum_i T_i) (\sum_i \rho_i)}{\sum_i (1 + R_i - T_i) \rho_i} \quad , \quad (47)$$

or in a slightly different form by replacing ρ_i by v_i the group velocity of electrons in channel i .

Problem 7 Derive (47) from (46).

Solve for the more general case in which the left-hand and right-hand leads are not identical, e.g. by having different numbers of open channels below the Fermi energy.

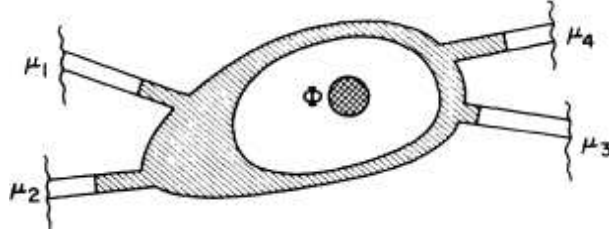


Figure 24: Disordered normal conductor with 4 terminals connected via perfect leads (unshaded) to 4 reservoirs at chemical potentials $\mu_1, \mu_2, \mu_3, \mu_4$. An Aharonov–Bohm flux may be applied through the hole in the sample.

5.3 Multiple leads

Inspection of figures 17 and 18 will show that we have missed an important feature of these experiments; namely that the potential may be measured between different contacts from those used for the current. This is, of course, an essential feature of Hall effect measurements. We first note the notation for the generalised resistance (32) which relates the potential difference between contacts i and j to the current flowing between contacts m and n . We will also consider purely 1-dimensional leads. As long as we are dealing with cases for which the voltage and current contacts are different we may ignore the considerations of the previous sections and write simply

$$I_i = 2 \frac{e}{h} \left[(1 - R_{ii}) \mu_i - \sum_{j \neq i} T_{ij} \mu_j \right] \quad (48)$$

where the first term is the net current in lead i due to the incident and reflected contributions from that lead, whereas the later terms are the contributions from the current transmitted from the other leads. This may be rewritten in matrix form as

$$\mathbf{I} = 2\frac{e^2}{h}\mathbf{G}'\mathbf{V} \quad , \quad (49)$$

where \mathbf{I} and \mathbf{V} are column vectors, \mathbf{G}' is a 4×4 matrix and we have used the relation $\mu_i = eV_i$. Current conservation demands that the sum of the elements of each column of \mathbf{G}' should be zero. Also, if all the μ 's are equal the current should be zero; hence the sum of the elements of each row should be zero. This also implies that the matrix \mathbf{G}' is singular. In the absence of a magnetic field time-reversal symmetry demands that $T_{ij} = T_{ji}$. This is relaxed to $T_{ij}(\Phi) = T_{ji}(-\Phi)$ in the presence of a magnetic field. Thus the 4×4 matrix \mathbf{G}' only has 8 (6) independent matrix elements with (without) a magnetic field.

In a typical experiment we expect no current to flow between the voltage contacts. Hence to calculate $\mathcal{R}_{12,34}$ we have to set $I_2 = -I_1$ and $I_3 = I_4 = 0$ and solve (49) for $V_3 - V_4$.

Problem 8 Derive an expression for $\mathcal{R}_{12,34}$ in terms of the transmission coefficients T_{ij} in the absence of a magnetic field. Note that we can't simply invert \mathbf{G}' ; it's singular.

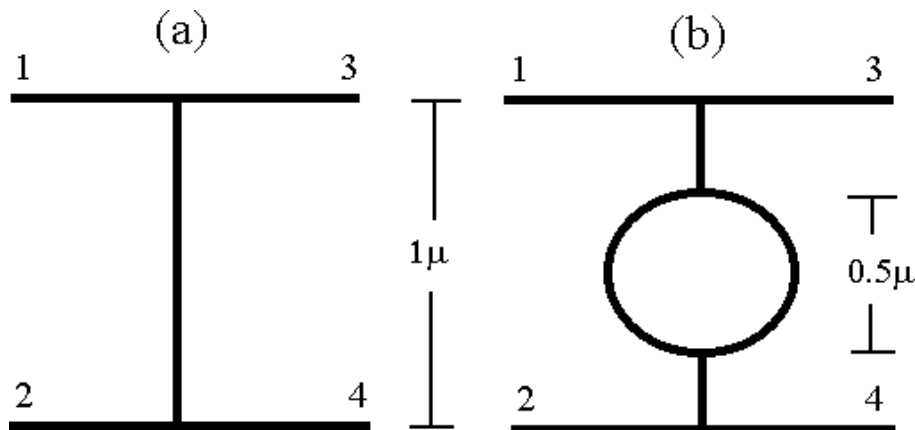


Figure 25:

Problem 9 Two samples of wires of submicron dimensions are prepared as in figure 25. Explain (quantitatively where possible) the behaviour you would expect for the temperature and magnetic field dependence of $R_{12,34}$ and $R_{13,24}$ in these samples.

A third sample is prepared in which many rings, like figure (b) and nominally all of the same diameter, are arranged in parallel between the two horizontal wires. What differences would you expect to observe in this sample as compared to sample (b)?

References

Abrahams E, Anderson P W, Licciardello D C & Ramakrishnan T V 1979 *Phys. Rev. Lett.* **42**, 673.

Aharonov Y & Bohm D 1959 *Phys. Rev.* **115**, 485.

Altshuler B L, Aronov A G & Spivak B Z 1981 *Sov. Phys. JETP Lett.* **33**, 94.

- Anderson P W 1958 *Phys. Rev.* **109**, 1492.
- Bergmann G 1984 *Phys. Rep.* **107**, 1.
- Büttiker M 1986 *Phys. Rev. Lett.* **57**, 1761.
- Büttiker M, Imry Y, Landauer R & Pinhas S 1985 *Phys. Rev.* **B31**, 6207.
- Efros A L & Shklovskii, B I 1975 *J. Phys. C: Solid State Physics* **8**, L49.
- Hewson A C 1993 *The Kondo Problem to Heavy Fermions* Cambridge University Press New York.
- Ioffe A F & Regel A R 1960 *Prog. Semicond.* **4**, 237.
- Kondo J 1969 *Solid State Physics* **23**, 184.
- Legendijk A 1994 in J. C Dainty, ed., ‘Current Trends in Optics’ Academic Press London pp. 51–68.
- Lambert C J 1991 *J. Phys. Condensed Matter* **3**, 6579.
- Landauer R 1970 *Phil. Mag.* **21**, 863.
- Lindelof P E, Nørregaard J & Hanberg J 1986 *Phys. Scr.* **T14**, 17.
- MacKinnon A & Kramer B 1983a *Z. Physik* **B53**, 1.
- MacKinnon A & Kramer B 1983b in G Landwehr, ed., ‘High Magnetic Fields in Semiconductors’ Vol. 177 of *Lecture Notes in Physics* Springer Berlin, Heidelberg p. 74.
- Mott N F 1968 *J. Non-Cryst. Sol.* **1**, 1.
- Mott N F & Davis E A 1979 *Electronic Processes in Non-Crystalline Materials* 2 edn Clarendon Press Oxford.
- Pepper M 1978 in L. R Friedman & D. P Tunstall, eds, ‘The Metal Non–Metal Transition in Disordered Solids’ SUSSP Edinburgh p. 285.
- Sharvin, D Y & Sharvin Y V 1981 *Sov. Phys. JETP Lett.* **34**, 272.
- Thouless D J 1974 *Phys. Rep.* **13**, 93.
- Umbach C P, Santhanam P, Van Haesendonck C & Webb R A 1987 *Appl. Phys. Lett.* **50**, 1289.
- van Wees B J, van Houten H, Beenakker C W J, Williamson J G, Kouwenhoven L P, van der Marel D & Foxon C T 1988 *Phys. Rev. Lett.* **60**, 848.
- Wharam D A, Thornton T J, Newbury R, Pepper M, Ahmed H, Frost J E F, Hasko D G, Peacock D C, Ritchie D A & Jones G A C 1988 *J. Phys. C: Solid State Physics* **21**, L209.