11 / Atom to transistor

In Chapter one, I used the generic structure shown in Fig.11.1.1a to focus and motivate this book. We spent Chapters 2 through 7 understanding how to write down a Hamiltonian matrix $[H_0]$ for the active region of the transistor structure whose eigenvalues describe the allowed energy levels (see Fig.11.1.1b). In Chapter 8, I introduced the broadening $[\Gamma_1]$ and $[\Gamma_2]$ arising from the connection to the source and drain contacts. In Chapter 9, I introduced the concepts needed to model the flow of electrons, ***neglecting*** phase-breaking processes. while in Chapter 10 we discussed the nature and meaning of phase-breaking processes, and how the resulting inflow and outflow of electrons is incorporated into a transport model. We now have the full "machinery" needed to describe dissipative quantum transport within the self-consistent field model (discussed in Chapter 3) which treats each electron as an independent particle moving in an average potential U due to the other electrons. I should mention, however, that this independent electron model misses what are referred to as "strong correlation effects"(see Prologue, Section 1.5) which are still poorly understood. To what extent such effects can be incorporated into this model remains to be explored (see Appendix, Section A.4).

My purpose in this Chapter is to summarize the machinery we have developed (Section 11.1) and illustrate how it is applied to concrete problems. I believe these examples will be useful as a starting point for readers who wish to use it to solve other problems of their own. At the same time, I have chosen these examples in order to illustrate conceptual issues that are of great importance in understanding the nature of electrical resistance on an atomic scale (Section 11.2-11.4).

11.1. Quantum transport equations

Fig.I0.1.1a. The generic transistor structure we used in the introduction to motivate this book.

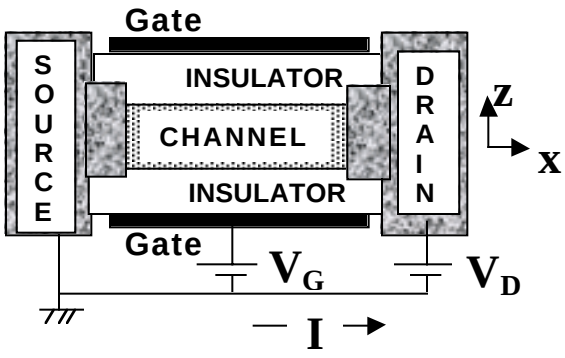
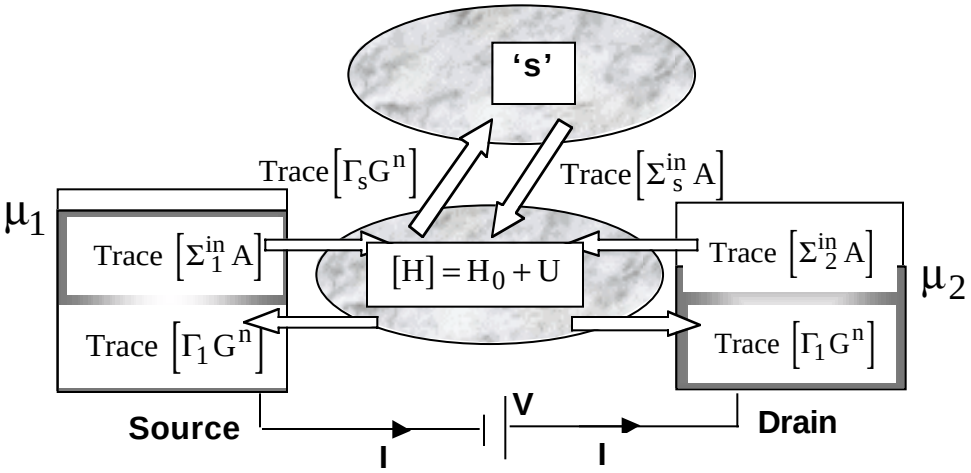


Fig.11.1.1b. Inflow and outflow of electrons for the generic structure in Fig.11.1a .



Let me quickly summarize the general model for dissipative quantum transport that we have discussed.

$$G^n = G\,\Sigma^{in}\,G^+ \tag{11.1.1}$$

$$G = [EI - H_0 - U - \Sigma]^{-1} \tag{11.1.2}$$

$$A = i[G - G^+] \;,\; \Gamma = i[\Sigma - \Sigma^+] \tag{11.1.3}$$

where $\Sigma^{in} = \Sigma_1^{in} + \Sigma_2^{in} + \Sigma_s^{in}$

$$\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_s \tag{11.1.4}$$

These equations can be used to calculate the correlation function G^n and hence the density matrix ρ whose diagonal elements give us the electron density.

$$\rho = \int dE\, G^n(E)/2\pi \tag{11.1.5}$$

The current (per spin) at any terminal 'i' can be calculated from

$$I_i = (q/\hbar) \int_{-\infty}^{+\infty} dE\ \tilde{I}_i(E)/2\pi \tag{11.1.6}$$

$$\text{with} \quad \tilde{I}_i = \text{Trace}\,[\Sigma_i^{in} A] - \text{Trace}\,[\Gamma_i G^n] \tag{11.1.7}$$

which is depicted in Fig.11.1b in terms of an inflow ($\Sigma_i^{in} A$) and an outflow ($\Gamma_i G^n$).

Input parameters: To use these equations, we need a channel Hamiltonian [H_0] and the inscattering [Σ^{in}] and broadening [Γ] functions. For the two contacts, these are related:

$$\Sigma_1^{in} = \Gamma_1 f_1 \qquad \text{and} \qquad \Sigma_2^{in} = \Gamma_2 f_2 \tag{11.1.8}$$

and the broadening / self-energy for each contact can be determined from a knowledge of the surface spectral function (a) / surface Green's function (g) of the contact and the matrices [τ] describing the channel contact coupling:

$$\Gamma = \tau a \tau^+ \quad \text{and} \quad \Sigma = \tau g \tau^+ \tag{11.1.9}$$

For all the numerical results presented in this chapter we will use the simple one-band effective mass (equal to the free electron mass) model, both for the channel Hamiltonian [H_0] and for the contact self-energy functions Σ_1, Σ_2.

For the scattering "terminal", unlike the contacts, there is no simple connection between Σ_s^{in} and Σ_s (or Γ_s). If the scattering process is essentially elastic ($E \approx E \pm \hbar\omega$), then (see Eqs.(10.3.13) – (10.3.15))

$$\Sigma_s^{in}(E) = D_0 \bullet G^n(E)\ ,$$
$$\Gamma_s(E) = D_0 \bullet A(E) \quad \text{and} \quad \Sigma_s(E) = D_0 \bullet G(E) \tag{11.1.10}$$

The phonon correlation function D_0 is generally a matrix (with the $D_0 \bullet G$ representing an element by element multiplication) and can be calculated from a knowledge of the phonon modes and the associated deformation potentials. But in our examples we will simply treat it as a scalar parameter. For general inelastic scattering processes (see Eqs.(10.3.9), (10.3.12))

$$\Sigma_s^{in}(E) = \int_0^\infty d(\hbar\omega) \begin{pmatrix} D^{em}(\hbar\omega) \bullet G^n(E+\hbar\omega) \\ + \ D^{ab}(\hbar\omega) \bullet G^n(E-\hbar\omega) \end{pmatrix}$$

$$\Gamma_s(E) = \int d(\hbar\omega) \begin{pmatrix} D^{em}(\hbar\omega) \bullet \left[G^p(E-\hbar\omega) + G^n(E+\hbar\omega)\right] \\ + \ D^{ab}(\hbar\omega) \bullet \left[G^n(E-\hbar\omega) + G^p(E+\hbar\omega)\right] \end{pmatrix}$$

but for our examples we will consider phonons with a single frequency ω_0

$$\begin{aligned} \Sigma_s^{in}(E) &= D_0^{em} \bullet G^n(E+\hbar\omega_0) + D_0^{ab} \bullet G^n(E-\hbar\omega_0) \\ \Gamma_s(E) &= D_0^{em} \bullet \left[G^p(E-\hbar\omega_0) + G^n(E+\hbar\omega_0)\right] \\ &\quad + \ D_0^{ab} \bullet \left[G^n(E-\hbar\omega_0) + G^p(E+\hbar\omega_0)\right] \end{aligned} \tag{11.1.11}$$

treating D_0^{em} and D_0^{ab} as scalar parameters and ignoring the Hermitian part of $\Sigma_s(E)$ which is given by the Hilbert transform of $\Gamma_s(E)$. Note that the inscattering and broadening functions in Eqs.(11.1.10) and (11.1.11) depend on the correlation functions, unlike the coherent case. This complicates the solution of the transport equations (Eqs.(11.1.1) – (11.1.4)), requiring in general an iterative self-consistent solution.

Diffusion equation: If the "phonon spectral function" D_0 is just a constant times an identity matrix, then it follows from Eq.(11.1.10) that Σ_s^{in} is a diagonal matrix with

$$\Sigma_s^{in}(r,r;E) = D_0\, G^n(r,r;E)$$

For long conductors one can neglect the contacts and write Eq.(11.1.1) as $G^n = G\,\Sigma_s^{in}\,G^+$, so that the diagonal elements of the correlation function which can be identified with the electron density, $n(r;E) = G^n(r,r;E)/2\pi$ obey the equation

$$n(r;E) = \sum_{r'} D_0 \left|G(r,r';E)\right|^2 n(r';E)$$

It can be shown that if the Green's function varies slowly in space then this equation reduces to the diffusion equation: $\nabla^2 n(r;E) = 0$ [Ref.A.2c].

Transmission: In Chapter 9 we saw that the concept of transmission is a very useful one and developed an expression for the current in terms of the transmission function (see Eq.(8.9)). For quantum transport with dissipation, the concept is still useful as a qualitative heuristic tool, but in general it is not possible to write down a simple quantitative expression for the current in terms of the transmission function, because there is no simple connection between Σ_s^{in} and Γ_s, unlike the contacts where $\Sigma_i^{in} = \Gamma_i f_i$. It is more convenient to calculate the current directly from Eq.(11.1.6) instead. We can define an effective transmission by comparing Eq.(11.1.6) with Eq.(9.9)

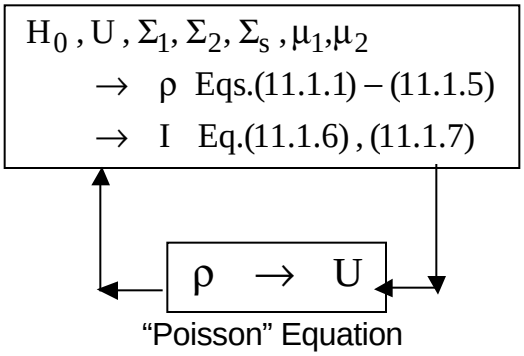
$$\bar{T}_{eff}(E) = \frac{\tilde{I}_i(E)}{f_1(E) - f_2(E)} = \frac{\text{Trace}\,[\Sigma_i^{in} A] - \text{Trace}\,[\Gamma_i G^n]}{f_1 - f_2} \qquad (11.1.12)$$

which can be a useful parameter to compare with coherent transport.

Self-consistent calculation: Finally I would like to note that in general it is necessary to perform a self-consistent solution of the transport equations with the "Poisson" equation which accounts for electron-electron interactions through a potential $U(\vec{r})$ (see Fig.11.1.2). We write Poisson within quotes as a reminder that this part of the problem could include corrections for correlation effects (see Chapter 3) in addition to standard electrostatics. This aspect is commonly ignored (as we do in Sections 11.2, 11.3) when calculating "linear response" for bias voltages that are small compared to the thermal energy k_BT and/or the energy scale on which the density of states changes significantly. The current is then independent of the precise spatial profile $U(\vec{r})$ arising from the applied drain voltage. But if we are interested in the shape of the current-voltage (I-V) characteristics over a wide range of

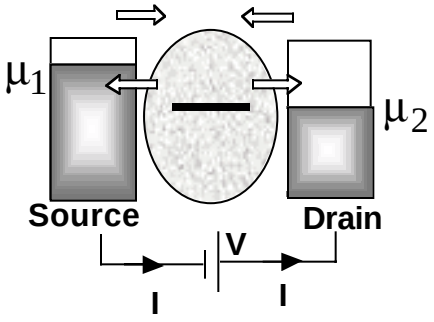
bias values as we are, for example, when calculating the ON-current of a transistor (Section 11.3), then the potential profile is of crucial importance as explained in the Prologue (see Section 1.4).

Fig.11.1.2. In general, the transport problem has to be solved self-consistently with the "Poisson" equation, which accounts for electron-electron interactions through the potential U.



A simple one-level example: In Chapter one, we went through an example with just one level so that the electron density and current could all be calculated from a rate equation with a simple model for broadening. I then indicated that in general we need a matrix version of this "scalar model" and that is what the rest of the book is about (see Fig.1.6.4).

Fig.11.1.3



Now that we have the full "matrix model" we have the machinery to do elaborate calculations as we will illustrate in the following calculations. But before getting into such details, it is instructive to specialize to a one-level system so that all the matrices reduce to pure numbers and the results are easily worked out analytically. From Eqs.(11.1.1) – (11.1.4),

$$G(E) = (E - \varepsilon + (i\Gamma/2))^{-1}$$

$$A(E) = \frac{\Gamma}{(E-\varepsilon)^2 + (\Gamma/2)^2} \quad , \quad \Gamma = \Gamma_1 + \Gamma_2 + D_0 A$$

$$G^n(E) = \Sigma^{in}(E)\, A(E)/\Gamma \quad , \quad \Sigma^{in} = \Gamma_1 f_1 + \Gamma_2 f_2 + D_0 G^n$$

Hence, $\frac{G^n}{A} = \frac{\Gamma_1 f_1 + \Gamma_2 f_2 + D_0 G^n}{\Gamma_1 + \Gamma_2 + D_0 A}$ so that $\frac{G^n}{A} = \frac{\Gamma_1 f_1 + \Gamma_2 f_2}{\Gamma_1 + \Gamma_2}$

independent of the scattering strength D_0. From Eqs.(11.1.6) – (11.1.7),

$$\tilde{I}_s = D_0 G^n A - D_0 A G^n = 0$$

and $$\tilde{I}_1 = \Gamma_1 A\,(f_1 - (G^n/A)) = \frac{\Gamma_1\Gamma_2 A}{\Gamma_1+\Gamma_2}(f_1 - f_2) = -\tilde{I}_2$$

$$\text{so that } I = \frac{q}{h}\int dE\, \frac{\Gamma_1\Gamma_2 A}{\Gamma_1+\Gamma_2}(f_1 - f_2)$$

showing that elastic phase-breaking of this sort in a one-level system has no effect on the current which is independent of D_0 .

11.2. Physics of Ohm's law

My objective in this Section is to show how the general quantum transport equations can be used to model conductors with phase-breaking processes. At the same time these examples will help illustrate how an ultrashort ballistic conductor evolves into a familiar macroscopic one obeying Ohm's law which states that the conductance is directly proportional to the cross-sectional area, S and inversely to the length, L. In the introductory chapter itself I noted that for a ballistic conductor it is easy to see why the conductance should increase with area (S) using elementary arguments. Now that we have discussed the concept of subbands (see Chapter 6) we can make the argument more precise. The conductance of a conductor increases with cross-sectional area because the number of subbands available for conduction increases and for large conductors this number is directly proportional to S.

But why should the conductance decrease with length (L)? Indeed a ballistic conductor without scattering has a conductance that is independent of its length. But for long conductors with scattering, the conductance decreases because the average transmission probability of electrons from the source to the drain decreases with the length of the conductor. We have seen in Chapter 9 that the conductance is proportional to the total transmission $\overline{T}$ at the Fermi energy (see Eq.(9.2.2)), which can be expressed as the product of the number of modes, M and the average transmission probability per mode, T:

$$G = (2q^2/h)\overline{T} = (2q^2/h)\,M\,T \qquad \textit{(Landauer formula)} \qquad (11.2.1)$$

For large conductors, M ~ S and T ~ 1 / L, leading to Ohm's law: G ~ S / L. And that brings us to the question: Why does the transmission probability decrease inversely with length?

Classical transport:

If we think of the electrons as classical particles, then it is easy to see why the transmission probability T ~ 1 / L. Consider a conductor consisting of two sections in series as shown in Fig.11.2.1. The first section has a transmission probability T_1 while the second has a transmission probability of T_2. What is the probability T that an electron will transmit through both? It is tempting to say that the answer is obviously $T = T_1 T_2$, but that is wrong. That is the probability that the electron will get through both sections in its first attempt. But an electron turned back from section 2 on its first attempt has a probability of $T_1T_2R_1R_2$ of getting through on after two reflections as shown in the figure ($R_1 = 1 - T_1$ and $R_2 = 1 - T_2$).

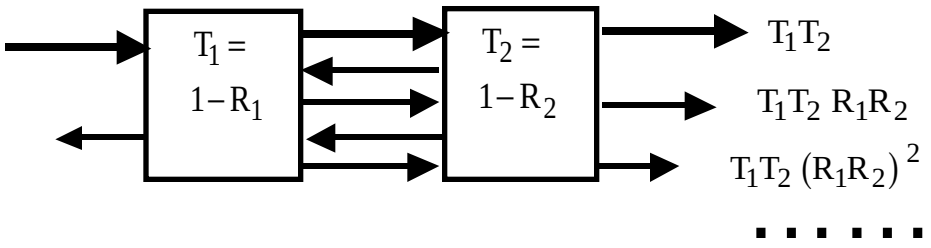


Fig.11.2.1. Classical "addition" of the transmission probabilities for two successive sections to obtain an overall transmission probability.

We can sum up the probabilities for all the paths analytically to obtain

$$T = T_1T_2\left((R_1R_2)+(R_1R_2)^2+(R_1R_2)^3+\cdots\right)$$

$$= \frac{T_1T_2}{1-R_1R_2} = \frac{T_1T_2}{T_1+T_2-T_1T_2}$$

so that

$$\frac{1}{T} = \frac{1}{T_1}+\frac{1}{T_2}-1 \qquad (11.2.2)$$

This relation tells us the resulting transmission T if we cascade two sections with transmission T_1 and T_2 respectively. From this relation we can deduce a general expression for the transmission, T(L) of a section of length L by asking what function will satisfy the relation

$$\frac{1}{T(L_1+L_2)} = \frac{1}{T(L_1)}+\frac{1}{T(L_2)}-1\ ?$$

It is easy to check that the following function fits the bill

$$T = \frac{L}{L+\Lambda} \qquad (11.2.3)$$

where Λ is a constant of the order of a mean free path, representing the length for which the transmission probability is 0.5.

Eq.(11.2.3) represents the transmission probability for classical particles as a function of the length of a conductor. Combining with Eq.(11.2.1) we obtain

$$G = (2q^2M/h)\frac{L}{L+\Lambda}$$

so that the resistance can be written as a constant interface resistance in series with a "device" resistance that increases linearly with length as required by Ohm's law:

$$\frac{1}{G} = \underbrace{\frac{h}{2q^2M}}_{\textit{interface resistance}} + \underbrace{\frac{h}{2q^2M}\frac{L}{\Lambda}}_{\textit{"device" resistance}} \qquad (11.2.4)$$

Eq.(11.2.4) suggests that the device conductance itself should be written as

$$G = \frac{2q^2M}{h}\frac{T}{1-T} \qquad (11.2.5)$$

which was the original form advocated by Landauer. Eq.(11.2.5) yields a resistance of zero as one might expect for a ballistic conductor (with T = 1), while Eq.(11.2.1) yields a non-zero resistance whose physical meaning caused extensive debate in the 1980's.

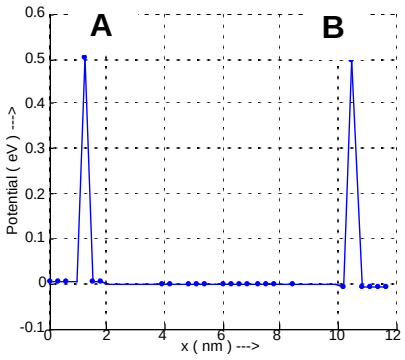
$$1/G = (h/2q^2M) \qquad 1/G = 0$$

Eq.(11.2.1) ***Eq.(11.2.5)***

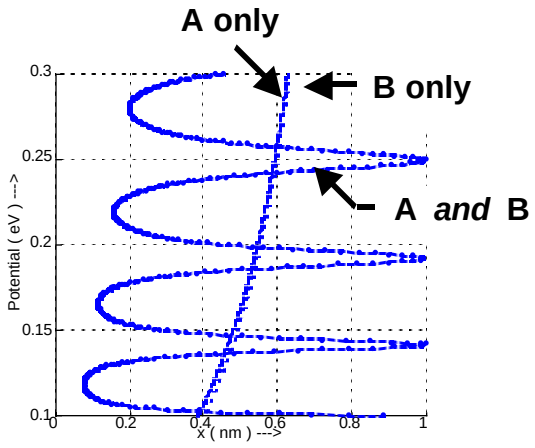
This non-zero resistance is now believed to represent the resistance associated with the interfaces between a low dimensional conductor with 'M' subbands and two large contacts with a very large number of subbands. In view of the fact that two-terminal measurements measure the total resistance rather than the "device" resistance, the present trend is to use Eq.(11.2.1) with the understanding that it includes the interface resistance along with the device resistance. Four-terminal measurements, on the other hand are typically interpreted on the basis of the multiterminal Buttiker formula discussed in Chapter 8.

Coherent transport (one subband):

We have seen that if we view electrons as classical particles, we do recover Ohm's law together with a constant interface resistance in series. But is this true if we treat electrons as quantum mechanical particles obeying the Schrodinger wave equation? The answer is no! This is easy to check numerically if we calculate the transmission through a device with one scatterer (A or B) and a device with two scatterers (A and B) as shown below.



(a) Device with two scatterers, A and B.



(b) Transmission versus energy with one scatterer only (A or B) and with both scatterers (A and B).

Fig.11.2.2. (a) A short device with two scatterers, A and B. A discrete lattice with 40 sites separated by a = 3 A was used in the calculation along with an effective mass equal to the free electron mass, m. A small bias potential was assumed varying linearly from +5 meV to –5 meV across the device. Each scatterer is represented by a potential of 0.5 eV at one lattice site.
(b) Transmission versus energy calculated with one scatterer only (A or B) and with both scatterers (A and B).

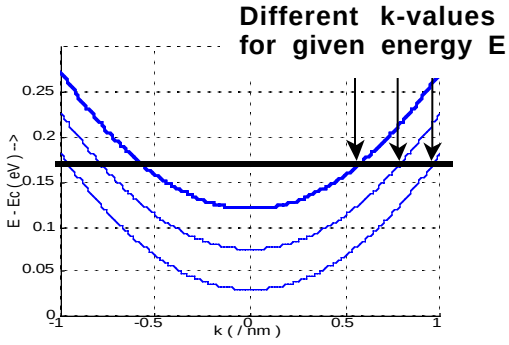
The transmission was calculated using the equations stated at the beginning of Section 11.2. with the phase-breaking terms (Σ_s and Σ_s^{in}) set equal to zero. Since we are dealing with coherent transport, we could calculate the transmission directly from Eq.(8.8). But it is better to use Eq.(11.2.7) since it is applicable to non-coherent transport and can be used in our later examples as well.

The important message from the example shown in Fig.11.2.2 is that the quantum transmission through two scatterers does not follow a simple rule like the one (Eq.(11.2.2)) we obtained for the transmission for classical particles. The basis reason is the interference between the two scatterers. If they are spaced by half a wavelength the reflections from the two scatterers interfere constructively, leading to a dip in the transmission. But if they are spaced by a quarter of a wavelength, the reflections interfere destructively, leading to a peak in the transmission. This shows up as large oscillations in the transmission as a function of energy (which determines the de Broglie wavelength of the electrons). Clearly then if we cascade two sections, we do not expect the composite transmission to have an additive property as required by Ohm's law. Indeed depending on the location of the Fermi energy, the transmission through two scatterers could even ***exceed*** that through the individual scatterers. This could never happen with classical particles which cannot have a higher probability of getting through two sections than of getting through one. But this is a well-known phenomenon with waves due to interference effects: light transmits better into a lens if we put an extra anti-reflection coating on top. Similarly with electrons too one section could act as an anti-reflection coating for the next section, leading to greater transmission and hence a lower resistance for two sections than for one!

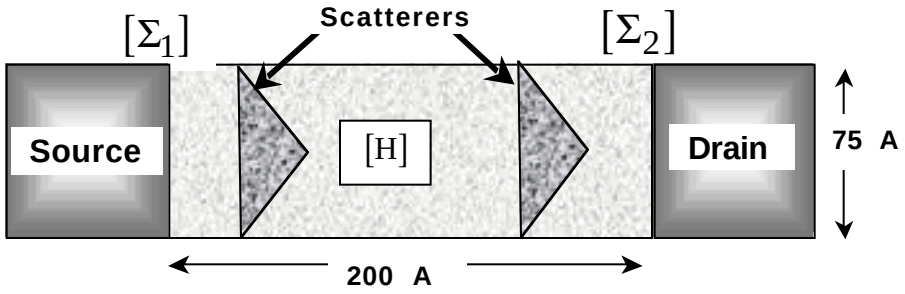
Coherent transport (multiple subbands):

One could argue that this is really an artifact of a one-dimensional model whereby electrons of a given energy have a single wavelength.

By contrast, in a multi-moded conductor, electrons of a given energy have many different values of 'k' and hence many different wavelengths 9in the longitudinal direction) , one for each mode, as shown.



As a result, we can expect interference effects to be diluted by the superposition of many oscillations with multiple wavelengths. This is indeed true. Fig.11.2.3 shows the transmission through a two-dimensional wire having a width of 75A and a length of 200A modeled with a discrete lattice of points spaced by 5A. Without any scatterer, the transmission at any energy E is equal to the number of propagating modes M(E), which increases in steps from four to six over the energy range shown. The transmission with one scatterer (with a transverse profile as shown below)



increases monotonically with energy and we can deduce a semiclassical transmission for two scatterers using Eq.(11.2.2) with $T = \overline{T} / M$ and noting that $T_2 \approx T_1$

$$\frac{M}{\overline{T}} = \frac{2M}{\overline{T}_1} - 1 \rightarrow \overline{T} = \frac{\overline{T}_1}{2 - (\overline{T}_1 / M)}$$

It is apparent that the quantum transmission through two scatterers fluctuates around this semiclassical result, the size of the fluctuation being of order one. Such fluctuations in the conductance of narrow wires as a function of the gate voltage (which shifts the Fermi energy relative to the levels) have been observed experimentally and are often referred to as universal conductance fluctuations. Fluctuations have also been observed as a function of the magnetic field which changes the effective wavelength at a given energy.

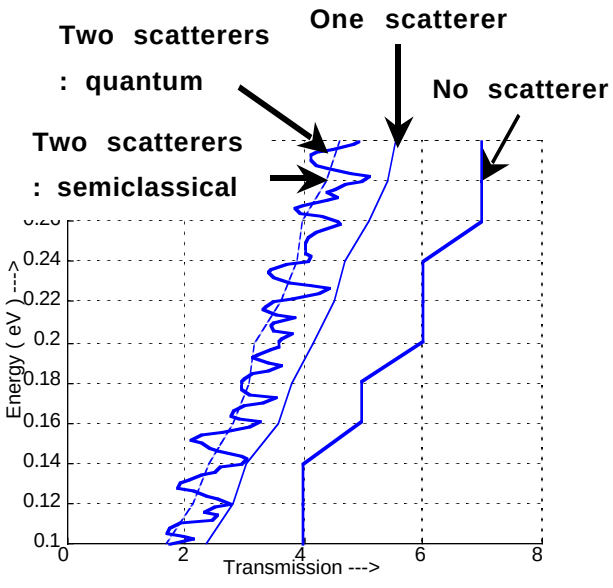


Fig.11.2.3. Transmission versus energy calculated for a wire with multiple subbands having no scatterer, one scatterer and two scatterers respectively. Each scatterer has a maximum potential of 0.25 eV at the center of the wire.

Since the size of the fluctuation is of order one, the reader might wonder what happens if the transmission falls below one. This will happen if the transmission probability T per mode M is smaller than 1/M, so that the total transmission $\overline{T}$ = MT is less than one. In Fig.11.2.4 we show the calculated transmission as the strength of the scatterers is increased from a maximum scattering potential of 0.25 eV to 5 eV. It is apparent that in the latter case the conductance shows large peaks separated by ranges of energy where the transmission becomes negligible indicating a ***strong localization*** of the electronic wavefunctions.

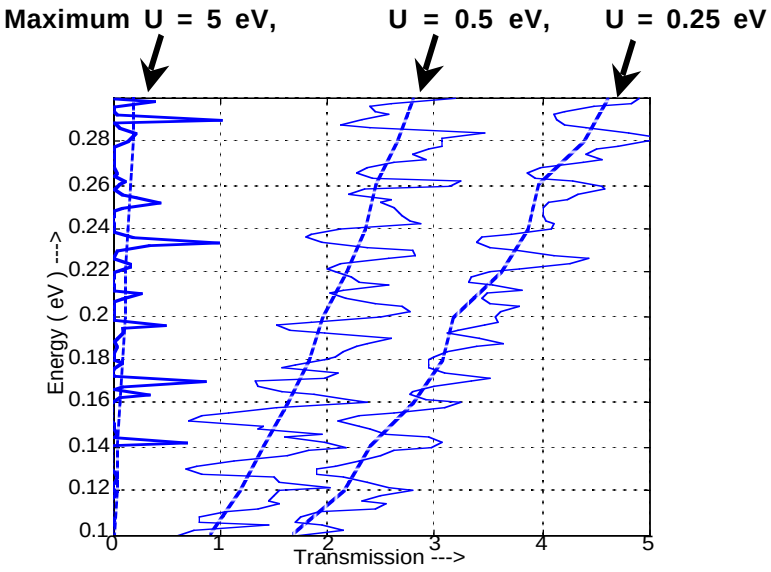


Fig.11.2.4. Transmission versus energy calculated for a wire with multiple subbands having two scatterers with a maximum potential of 5 eV, 0.5 eV and 0.25 eV respectively. The dashed lines show the semiclassical result for two scatterers deduced from the transmission through one scatterer.

By contrast, with weaker scattering potentials when the semiclassical transmission is larger than one, the quantum transmission shows fluctuations of order one around the semiclassical result. With multiple scatterers, the average quantum transmission turns out to be approximately one ***less*** than the semiclassical result. This is not at all evident from Fig.11.2.4 which only involves two scatterers. With a large number of independent scatterers, it turns out that the backscattering per mode is enhanced by (1/M) due to constructive interference leading to an equal decrease in the transmission per mode and hence to a decrease of one in the total transmission. A magnetic field destroys the constructive interference, causing an increase in the transmission which has been experimentally observed as a negative magnetoresistance (reduction in the resistance in a magnetic field) and is ascribed to this so-called ***weak localization*** effect. The basic phenomenon involving a coherent increase in backscattering has also been observed with electromagnetic waves in a number of different contexts unrelated to electron transport.

Quantum transport with dephasing

To summarize, as we make a wire longer the semiclassical transmission will decrease in accordance with Ohm's law and the quantum transmission will exhibit fluctuations of the order of one around the semiclassical result. If we make any wire long enough, there will come a length for which the semiclassical transmission will be less than one (see Eq.(11.2.3)):

$$\bar{T} \approx \frac{M\Lambda}{L+\Lambda} < 1 \text{ if } \quad L > M\Lambda$$

The quantum transmission for a wire longer than the localization length ($\sim M\Lambda$) will show large fluctuations characteristic of the strong localization regime. It would seem that even a copper wire if it is long enough will eventually enter this regime and cease to obey anything resembling Ohm's law! However, that is not what happens in real life. Why?

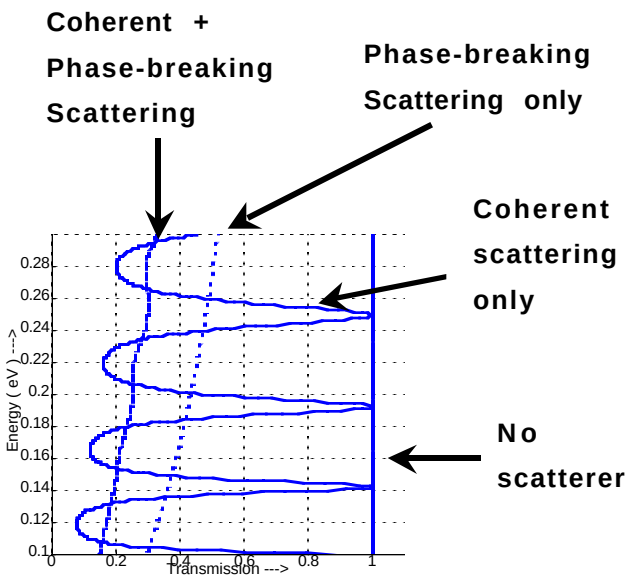


Fig.11.2.5. Transmission versus energy for a one-dimensional wire with different types of scattering as indicated.

The reason is that our observations are valid for phase-coherent conductors where we do not have significant phase-breaking processes to dilute the quantum interference effects. A wire will exhibit strong localization only if the localization length $M\Lambda$ is shorter than the phase-breaking length. Since this length is typically

quite short especially at room temperature, there is little chance of a copper wire (with its enormous number of modes, M) ever entering this regime. Fig.11.2.5 shows the effective transmission for a one-dimensional wire having two coherent scatterers, with and without phase-breaking scattering. It is apparent that the interference effects are effectively washed out by the presence of phase-breaking processes. Fig.11.2.6 shows that a one-dimensional wire with only phase-breaking scatterers leads to Ohm's law like behavior as a function of length. Of course in this limit a full quantum transport model is unnecessary. We could probably use a semiclassical model that neglects interference effects altogether and treats electrons as particles [11.2]. However, in general, we have both coherent and phase-breaking scattering and the quantum transport model described in Section 11.1 allows us to include both.

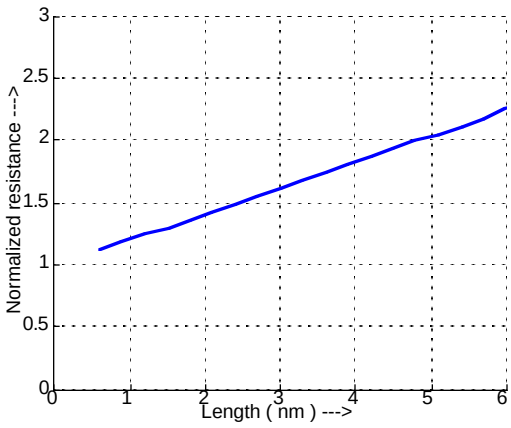


Fig.11.2.6. Normalized resistance (inverse transmission) as a function of length for a one dimensional wire with phase-breaking scattering only.

In these calculations we have assumed that the phase-breaking scatterers carry negligible energy away ($\hbar\omega \rightarrow 0$), so that we can use the simplified versions (cf. Eqs.(11.10))

$$\Sigma_s^{in}(E) = D_0\,[G^n(E)] \qquad \text{and} \qquad \Sigma_s(E) = D_0\,[G(E)] \tag{11.2.6}$$

to evaluate the self-energy and inscattering functions. We are also assuming the scattering to be diagonal in real space and uniformly distributed so that D_0 is a constant that we have set equal to 0.01 eV^2 in Fig.11.2.5 and to 0.05 eV^2 in Fig.11.2.6.

11.3. Where is the heat dissipated?

We have seen that with adequate amounts of phase-breaking scattering, the resistance of a conductor increases linearly with length (Fig.11.2.6) in accordance with Ohm's law. However, as the length tends to zero (ballistic conductor), the resistance tends to a constant representing the interface resistance associated with the interfaces between the low-dimensional conductor and the three-dimensional contacts. But where does the associated I^2R loss (or the Joule heating) occur?

The answer to this question depends on the nature of the scattering process. The point is that resistance comes from the loss of momentum associated with scattering while the associated Joule heating comes from the loss of energy. For example, in Figs.11.2.5 and 11.2.6 we have modeled the phase-breaking scattering as an elastic process, neglecting any associated energy loss. This means that in this model, no energy is dissipated inside the device at all. Nonetheless this elastic scattering does give rise to a resistance that obeys Ohm's law because of the associated loss of momentum, not the loss of phase. Indeed in small conductors, a significant fraction of the Joule heat I^2R associated with a resistor R could be dissipated in the contacts rather than in the conductor itself. There is concrete evidence that this is true, allowing experimentalists to pump far more current through small conductors than what would be needed to destroy them if all the heat were dissipated inside them.

The fact that elastic scattering is not associated with any energy loss can be seen by noting at the normalized current per unit energy $\tilde{I}_i(E)$ (see Eq.(11.1.6)) is identical at each of the two terminals (source and drain) as shown in Fig.11.3.1.

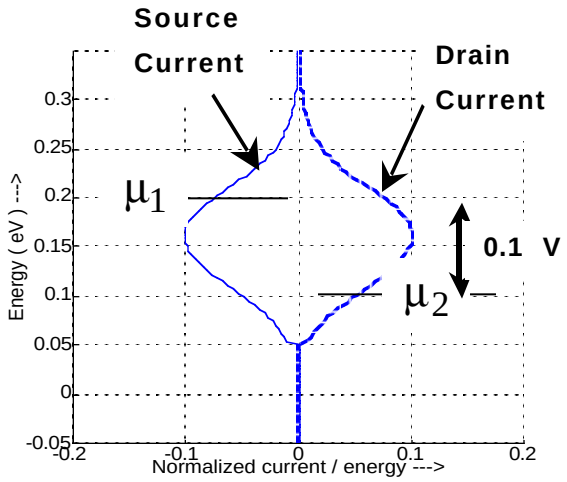


Fig.11.3.1. Normalized current per unit energy, $\tilde{I}_i$ (E) in a one-dimensional wire with phase-breaking elastic scattering.

The point is that the ***energy current*** at any terminal is given by (cf. Eq.(11.1.6))

$$I_{E,i} = \int_{-\infty}^{+\infty} dE \; E \, \tilde{I}_i(E) / 2\pi\hbar \qquad (11.3.1)$$

and if there is power (P_d) dissipated inside the device then it must be reflected as a difference in the energy currents at the source and drain terminals:

$$P_d = I_{E,\,drain} - I_{E,\,source} \qquad (11.3.2)$$

Since current conservation requires the current to be the same at the source and the drain, the energy currents can be different only if they are distributed differently as a function of energy. Clearly there is no power dissipated in the device shown in Fig.11.3.1, since the current has the same energy distribution at the source and drain. But if we model the same device assuming that the scatterers have an associated phonon energy of $\hbar\omega$ = 20 meV, the energy distribution of the current is different at the source and drain, showing that some fraction of the I^2R loss occurs inside the device. Electrons enter the source at a higher energy on the average than the energy at which they exit the drain.

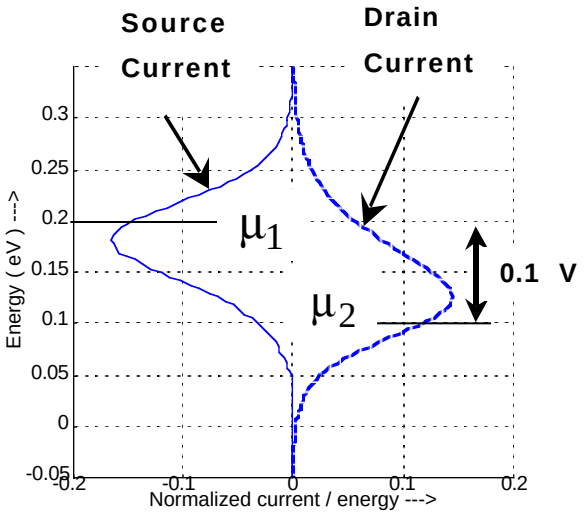


Fig.11.3.2. Normalized current per unit energy, $\tilde{I}_i(E)$ in a one-dimensional wire with inelastic scattering by phonons with energy $\hbar\omega$ = 20 meV (D_0 = 0.1 eV^2). The drain current flows at a lower energy than the source current due to the energy relaxation inside the device.

"Peltier" effect:

It is interesting to note that power need not be dissipated everywhere. In an inhomogeneous device there could be local regions where energy is absorbed by the electrons and the solid is locally ***cooled***. Consider, for example, a one-dimensional current with a potential step in the center that forces the current to flow at a higher energy at the drain terminal than at the source terminal (Fig.11.3.2). It is easy to see that the junction is cooled by the flow of current, which can be considered a microscopic version of the well-known Peltier effect where an electrical current cools one junction at the expense of another. Indeed if we had a potential barrier with an upstep followed by a downstep then the upstep would be cooled while the downstep would be heated. This aspect of the energy exchange is reversible and is proportional to the current, unlike the irreversible Joule heating which is proportional to the square of the current.

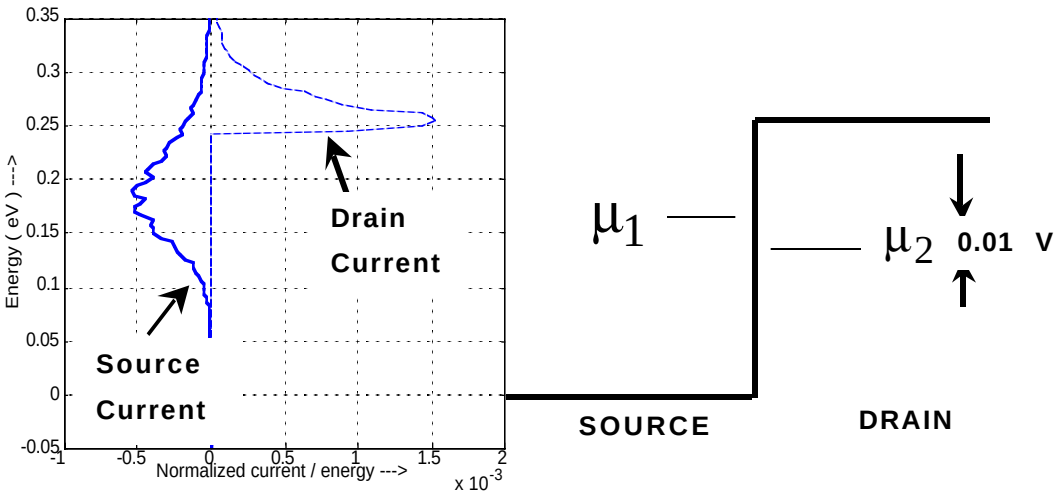


Fig.11.3.3. Normalized current per unit energy, $\tilde{I}_i(E)$ in a one-dimensional wire with inelastic scattering by phonons with energy $\hbar\omega$ = 20 meV (D_0 = 0.1 eV^2), having a potential step in the middle as shown. The drain current flows at a higher energy than the source current indicating that the device is cooled by the flow of current.

11.4. Where is the voltage drop?

In Section 11.1, I stressed the importance of doing quantum transport calculations self-consistently with the "Poisson" equation (see Fig.11.1.2) for the self-consistent potential U representing the effect of the other electrons. This is particularly important when calculating the current under a "large" applied voltage: the shape of the current-voltage characteristics can sometimes be significantly different depending on the potential profile (or the "voltage drop") inside the channel. For example, in determining the maximum (or the ON) current of a transistor it is important to know where the voltage drops.

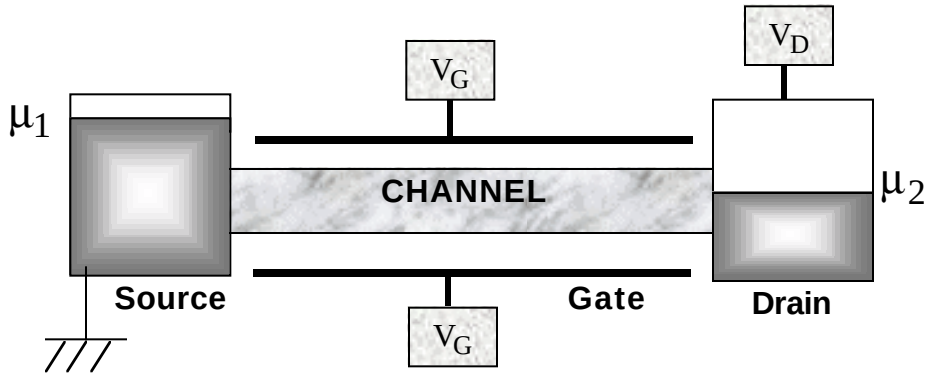


Fig.11.4.1. A nanotransistor consisting of a quantum wire channel surrounded by a coaxial gate which is used to induce electrons in the channel.

Consider a "nanotransistor" composed of a narrow quantum wire labeled the "channel" (see Fig.11.4.1) of radius 'a' surrounded by a coaxial gate of radius 'b' which is used to induce electrons in the channel as we discussed in Chapter 7. Assume that the electrons in the channel belong to a single subband with a parabolic dispersion relation

$$E = E_c + (\hbar^2 k^2 / 2m) \qquad (11.4.1)$$

At equilibrium, with $\mu_2 = \mu_1$ and low temperatures ($T \rightarrow 0$ K) the density of electrons (per unit length) in the channel can be written as (see Table 5.2.1 with an extra factor of two to account for two spins)

$$n_L = 2\sqrt{2m(\mu_1 - E_c)}/\pi\hbar \qquad (11.4.2)$$

If we make the gate voltage V_G more positive, it will induce more electrons in the channel, while if we make it more negative it will deplete the channel of electrons, in much the same way that we discussed in Section 7.3, except that in Chapter 7 we were talking primarily about a flat two-dimensional conductor, while now we are talking about a cylindrical one-dimensional conductor. In Chapter 6 we discussed only the equilibrium problem with $\mu_2 = \mu_1$. The problem I wish to discuss now is a non-equilibrium one. A voltage V_D is applied to the drain relative to the source making $\mu_2 = \mu_1 - qV_D$. What is the current, I? Formally we can calculate by following the self-consistent procedure depicted in Fig.11.1.2 and numerical results are shown in Fig.11.4.3. But first let us try to understand the physics in simple terms.

Ballistic nanotransistor: We will start with a ballistic transistor (no scattering) having perfect contacts. If the contacts are good and there is no scattering we would expect the low bias conductance to be equal to the conductance quantum times two (for spin).

$$I = (2q^2/h)\,V_D \qquad (11.4.3a)$$

The +k states are filled up from the left contact with an electrochemical potential μ_1 while the – k states are filled from the right contact with an electrochemical potential μ_2. In the energy range between μ_1 and μ_2 (plus a few k_BT on either side) the +k states are nearly filled and carry current, but this current is not balanced by the –k states since they are nearly empty. Since a 1-D wire carries a current of (2q/h) per unit energy, there is a net current given by (2q/h) times the energy range (μ_1-μ_2) which is equal to $(2q^2/h)V_D$ as stated above.

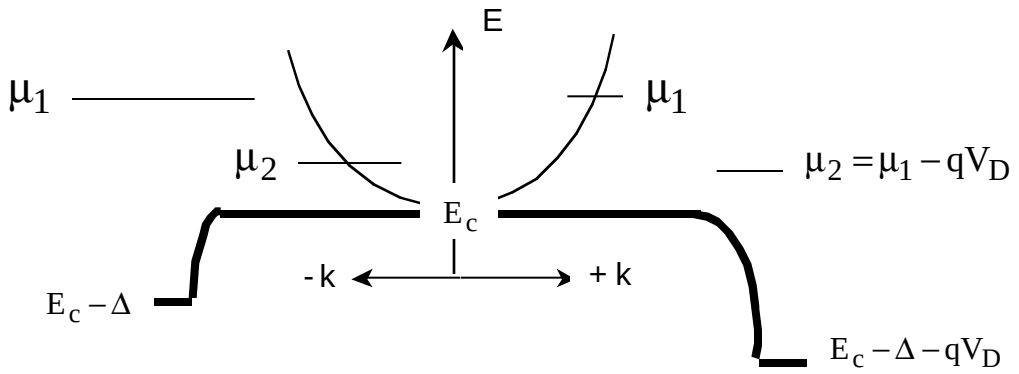


Fig.11.4.2. An applied voltage lowers the energy levels and the electrochemical potential in the drain region. The +k states are occupied from the source upto μ_1 while the –k states are occupied from the drain upto μ_2 causing a net current to flow as discussed in Chapter 6 (see Section 6.3).

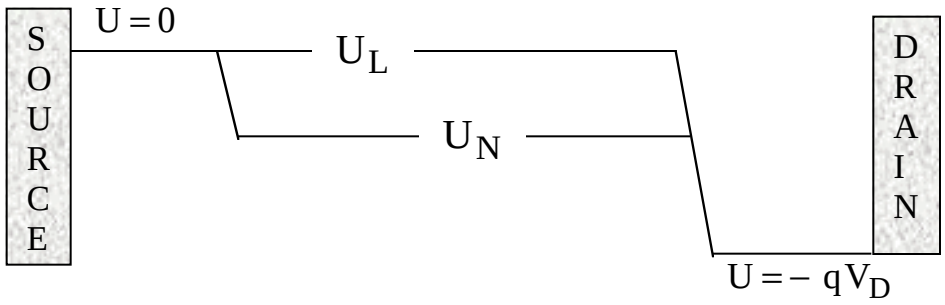


Fig.11.4.3. Same as Fig.11.4.2 except that the equilibrium conduction band profile has been *subtracted* off. This plot shows just the change U in the potential under an applied bias. Inside the channel, the two extreme possibilities are the Laplace potential U_L and the neutral potential U_N as explained in the text.

Once μ_2 has dropped below the bottom of the band (E_c), the current cannot increase any further and we expect the ON-current to be given by

$$I_{ON}^{(L)} = (2q^2/h)\ (\mu_1 - E_c) \qquad (11.4.3b)$$

Eqs.(11.4.3a, b) suggest that the current should increase linearly with voltage and then level off as the voltage approaches $(\mu_1 - E_c)$. It is indeed true that the current increases linearly and then saturates, but depending on the electrostatics, the ON-current could be much larger, upto ***four*** times as large as that given by Eq.(11.4.3b), which we have labeled with an 'L' to denote the Laplace limit. Let me explain what I mean by that and what the other limit is.

The result given in Eq.(11.4.3b) is based on the picture shown in Fig.11.4.2 which assumes that the only effect of the increasing drain voltage is to lower μ_2, ***while the energy levels in the channel remain fixed*** relative to the source. However, depending on the electrostatics, it is possible that the potential energy U in the channel would drop by some fraction of the drain potential $-qV_D$ thus lowering the bottom of the band and increasing the ON-current to (note that U is a negative quantity for positive drain voltages)

$$I_{ON} = (2q^2/h)\ (\mu_1 - E_c - U) \qquad (11.4.4)$$

(I am assuming that U remains less than Δ, see Fig.11.4.2). So in estimating the ON-current, the all important question is "How does the voltage drop?" The source is at zero, the drain is at $-qV_D$: What is the potential energy U inside the channel?

In general we determine the channel potential from a self-consistent solution of the electrostatics and the transport problems. For our present problem we can write the electrostatics in the form

$$U = U_L + (q^2/C_E)\,\delta n_L \qquad (11.4.5)$$

where C_E is the capacitance per unit length of a coaxial capacitor with inner and outer radii equal to 'a' and 'b' respectively.

$$C_E = 2\pi\varepsilon_r\varepsilon_0 / \ln(b/a) \qquad (11.4.6)$$

δn_L is the change in the electron density per unit length. and U_L is the solution to corresponding to $\delta n_L = 0$, sometimes called the Laplace solution which is shown in Fig.11.4.3. Since the potential is applied only to the drain and not the gate the Laplace potential has the shape shown in Fig.11.4.3. It is essentially equal to the

source potential throughout the channel and rapidly changes to the drain potential at the other end; how rapidly depends on the closeness of the gate to the channel.

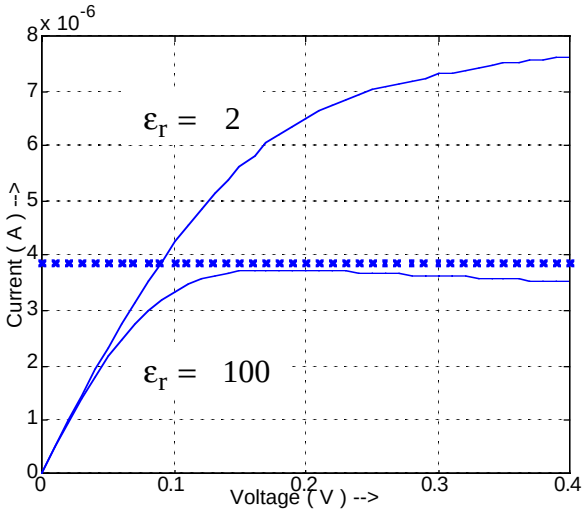


Fig.11.4.4. Current (I) vs. Drain voltage (V_D) for a ballistic quantum wire transistor, for different values of the capacitance C_E. Also shown with x's is the Laplace limit with U = U_L. (Courtesy of S. Salahuddin)

The actual potential inside the channel is close to the Laplace limit if the electrostatic capacitance C_E is large: the second term in Eq.(11.4.5) then is negligible. This is the case when we assume a very "high-K" dielectric with ε_r = 100 (see Eq.(11.4.6) with b set equal to '2a'). But when we use a smaller ε_r = 2, the current increases significantly.by nearly a factor of two.

If the capacitance is even smaller, then in principle we could be in the other limit where $\delta n_L \rightarrow 0$, but the second term is finite. We call this the neutral limit and the corresponding potential the neutral potential U_N (see Section 7.3). What is U_N? We can write the electron density in the ON-state as

$$[n_L]_{ON} = \sqrt{2m(\mu_1 - E_c - U)}/\pi\hbar) \qquad (11.4.7)$$

since the potential energy U (which is negative) moves the bottom of the band down to E_c + U, but we lose a factor of two because only half the states (having +k) are occupied. Subtracting Eq.(11.4.7) from (11.4.2) we write the change in the electron density as

$$\delta n_L = (\sqrt{2m}/\pi\hbar)\,(2\sqrt{\mu_1 - E_c} - \sqrt{\mu_1 - E_c - U}) \qquad (11.4.8)$$

Setting $\delta n_L = 0$, we obtain the neutral potential:

$$2\sqrt{\mu_1 - E_c} = \sqrt{\mu_1 - E_c - U_N}$$
$$\rightarrow \quad \mu_1 - E_c - U_N = 4\,(\mu_1 - E_c) \qquad (11.4.9)$$

which means that in the neutral limit the ON-current from Eq.(11.4.4) is four times what we expect in the Laplace limit (cf. Eq.(11.4.3b))

$$I_{ON}^{(N)} = (2q^2/h)\;(\mu_1 - E_c - U_N) = 4\; I_{ON}^{(L)} \qquad (11.4.10)$$

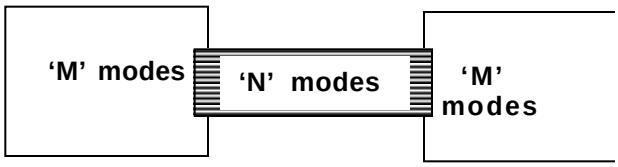
I should note that the neutral limit of the ON-current need not always be four times the Laplace limit. The factor of four is specific to the one-dimensional example considered here arising from the fact that the electron density is proportional to the square root of the energy (see Eq.(11.4.1)). For a two-dimensional sheet conductor, the electron density increase lenearly with energy and we can show that the neutral limit is two times the Laplace limit. The important point is that there is a Laplace limit and a neutral limit and the actual value could lie anywhere in between depending on the capacitance C_E.

Electrostatic boundary conditions: In Fig.11.4.3 we have shown the potential U approaching the asymptotic values of zero and - qV_D set by the external voltage in the source and drain regions respectively. It is common to assume that this statement will be true if we make these regions long enough. However, it is important to note that if the end regions are assumed ballistic then the potential may not reach the asymptotic values, no matter how long we make these regions.

Fig.11.4.5. Sketch of the spatial profile of the neutral potential: In the end regions it does not approach the correct asymptotic values of zero and -qV_D because a fraction of the density of states in these regions is "controlled" by the contact at the other end.

The reason is that in these conductive end regions the potential U will approach the neutral value U_N needed to make δn_L = 0. Consider the region near the source, for example. If the potential U were zero, δn_L would be negative because a fraction of the density of states in this region is now occupied according to the electrochemical potential μ_2 in the drain. This fraction is described by the partial spectral function [A_2] that we discussed in Chapter 9. To keep δn_L = 0, the neutral potential in this region takes on a negative value (Fig.11.4.5). This situation will not change if we simply make the end regions longer. As long as they are ballistic, it can be shown that there will be no change in the fraction of the density of states at one end that is controlled by the contact at the other end. Consequently the neutral potential in the source region will be less than the asymptotic value of zero and using a similar argument we can show that in the drain region it will be more than the asymptotic value of -qV_D.

The potential U will revert to the correct asymptotic values $\pm qV/2$ only if a negligible fraction of the density of states (or spectral function) at one end is controlled by the contact at the other end. This can happen if there is enough scattering within the device or if there is strong geometrical dilution at the contacts (M >> N).



This means that in modeling near ballistic devices without significant geometric dilution at the contacts we should not fix the potential at the two ends to the usual asymptotic values as we did in solving the capacitance problem (see Eq.(7.2.18)). One solution is to use a zero-field boundary condition for the Poisson equation and let the potential U develop self-consistently.

From a conceptual point of view, we could view the spatial profile of the neutral potential U_N (which may be different from the profile of the actual potential U) as an indicator of the spatial distribution of the resistance. The neutral potential across any ballistic region remains flat indicating zero resistance as we might intuitively expect. Fig.11.4.5 shows that only a fraction of the applied voltage V_D actually appears between the two ends of the device indicating that the resistance we calculate is only partly due to the channel and the rest should be associated with the interface between the narrow channel regions and the wide contact regions.

Nanotransistor with scattering: We expect the ON-current to be reduced by the presence of scattering, since the transmission is now less than one by a factor (see Eq.(11.2.3))

$$T = L/(L+\Lambda) \tag{11.4.11}$$

that depends on the length of the channel relative to a mean free path. In practice, however, the ON-current is higher that what Eq.(11.4.10) suggests, if the scattering processes are inelastic rather than elastic.

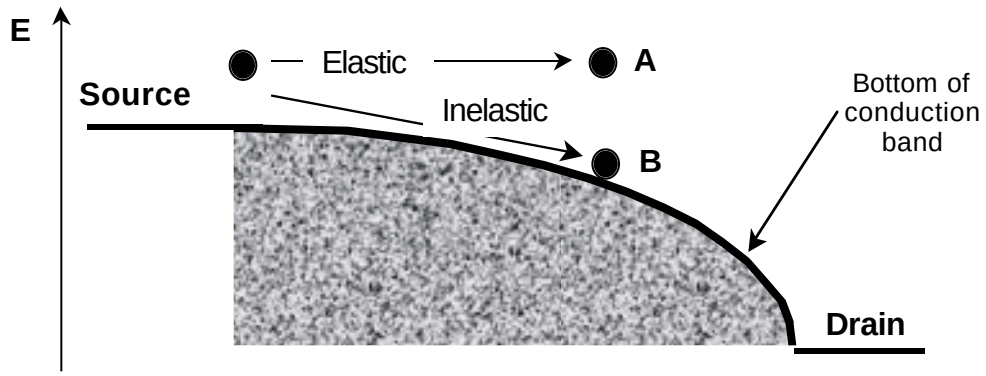


Fig.11.4.5. Once an electron has lost sufficient energy through inelastic processes (point B) it cannot be turned back towards the source and current is not reduced by further scattering. But if it has not lost enough energy (point A) then backscattering to the source can occur.

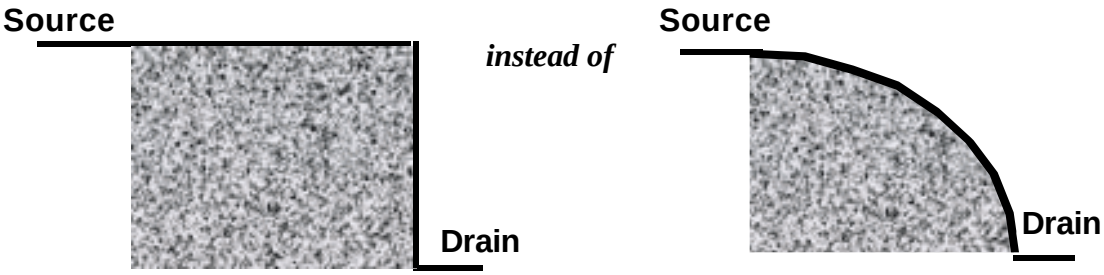
To understand why, we note that inelastic processes cause the electrons coming in from the source to lose energy as they propagate towards the drain (se Fig.11.4.5). Once they have lost sufficient energy (indicated by point B) they cannot easily be turned back towards the source any more, since there are no allowed states in the source at this lower energy. Consequently, the electron proceeds to the drain and the current is not reduced. But if the scattering processes are elastic then electrons do not relax in energy (indicated by point A) and can be turned back towards the source with a reduction in the current. This physics can be described approximately by replacing the device length L in Eq.(11.4.10) with the energy relaxation length L_i that an electron traverses before it loses a few k_BT of energy (large enough that coming back is near impossible)

$$T = L_i/(L_i + \Lambda) \qquad (11.4.11)$$

This can make the actual ON current much larger than what one might expect otherwise. Even purely elastic scattering causes a similar increase in the ON current in two-dimensional conductors since it relaxes the ***longitudinal*** (directed from the source to the drain) kinetic energy, although the total energy (longitudinal + transverse) remains the same. Indeed, commercial transistors for many years now

have shown ON currents that are within 50% of their ballistic value even though they may be over ten mean free paths in length. The reason is that the energy relaxation length L_i tends to be of the same order as the mean free path Λ, making the transmission probability approximately 0.5 regardless of the actual length of the channel.

My reason for bringing up this issue here is that this is another example of the importance of the self-consistent potential profile in determining the current at large bias voltages. For example, the effect just described would not arise if the potential profile looked like



since there would then be little room for energy relaxation. It seems reasonable to expect that the actual magnitude of the effect in a real device will depend on the potential profile that has to be calculated self-consistently as indicated in Fig.11.1.2. This is a point that is often overlooked because the transport block in Fig.11.1.2 is intellectually more demanding than the "Poisson" block and tends to overshadow it in our mind. So it is worth remembering that in many problems, like the ON current of a nanotransistor, the Poisson block could well represent the key "physics" making the transport block just a "detail"!

Exercises

E.11.1. (a) Consider a short 1-D channel with two scatterers, A and B modeled with a discrete lattice of 40 sites separated by a = 3 A. Assume a small bias potential varying linearly from +5 meV to –5 meV across the device. Each scatterer is represented by a potential of 0.5 eV at one lattice site. Calculate the transmission versus energy with one scatterer only (A or B) and with both scatterers (A and B) and compare with ***Fig.11.2.2***.
(b) Repeat including elastic phase-breaking scattering processes as indicated in Eq.(11.2.6) and compare with ***Fig.11.2.5.***
(c) Plot the inverse transmission (at a fixed energy of 0.1 eV) versus length for a 1-D wire with elastic phase-breaking scattering only and compare with ***Fig.11.2.6.***

E.11.2. (a) Consider a multimoded channel 75A wide and calculate the transmission through one scatterer and through two scatterers assuming each scatterer to be represented by a triangular potential (with a maximum of 0.25 eV at the center) in the transverse direction localized at one lattice plane in the longitudinal direction. Also, plot the semiclassical result obtained for two scatterers using the result for one scatterer. Compare with ***Fig.11.2.3***.
(b) Repeat for different strengths of the scattering potential and compare with ***Fig.11.2.4.***

E.11.3. (a) Consider a 1-D channel with phase-breaking elastic scattering and plot the current per unit energy as shown in ***Fig.11.3.1***.
(b) Repeat for a channel with inelastic scattering and compare with ***Fig.11.3.2.***
(c) Repeat for a channel with a potential step in the middle with inelastic scattering and compare with ***Fig.11.3.3.***

E.11.4. Calculate the current (I) versus drain voltage (V_D) self-consistently for the ballistic quantum wire nanotransistor (Fig.11.3.1) and compare with ***Fig.11.4.4.***

E.11.5. Tunneling in the presence of phase-breaking: Calculate the inverse transmission at low drain voltage at E = μ for a conductor having its equilibrium electrochemical potential μ located at 0.1 eV, with the conduction band edge in the contact at 0 eV and that in the channel at 0.5 eV as shown in the figure below:

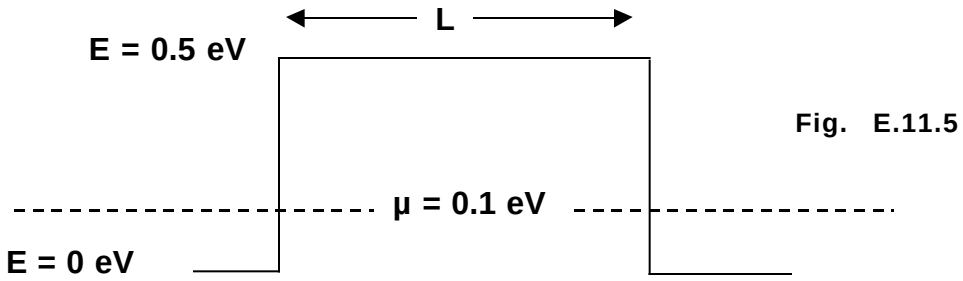
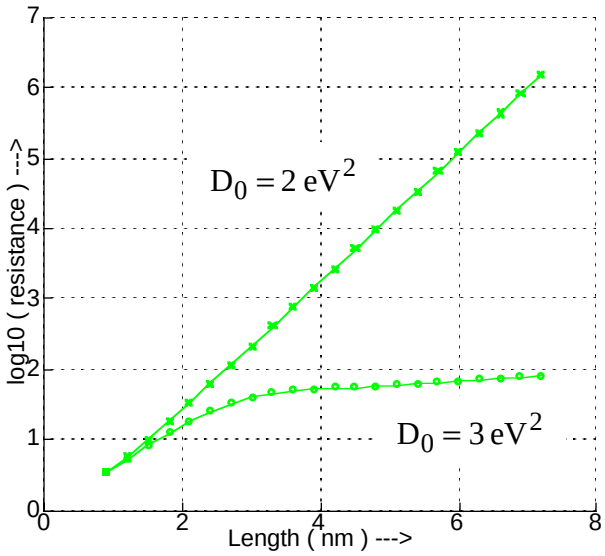


Fig. E.11.5

Transmission through the channel is by tunneling, so that the inverse transmission varies exponentially with the length 'L' of the barrier in the absence of phase-breaking processes. Plot the logarithm of the inverse transmission (normalized resistance) as a function of the length for 1 nm < L < 7 nm with two different scattering strengths, $D_0 = 2\,\mathrm{eV}^2$ and $3\,\mathrm{eV}^2$ respectively.

You should obtain a plot like this



Note that the expected exponential dependence of the resistance (linear dependence of the logarithm of the resistance) does not hold at higher scattering strengths. See for example, G. Neofotistos, R. Lake and S. Datta, "Inelastic Scattering Effects on Single Barrier Tunneling," Phys. Rev. B 43, Rapid Communications, 2242-2445 (1991).

