Lecture 3

Resistance: Ballistic to Diffusive

Contents

3.1 Introduction

We are now ready to use the general model introduced in Lecture 2. We will consider a simple problem, determining the resistance of 1D, 2D, and 3D resistors beginning with very short (ballistic) resistors, then treating conventional, diffusive resistors, and finally treating the entire ballistic to diffusive spectrum. As sketched in Fig. 3.1, in a 1D resistor (a "nanowire") electrons are free to move in only one dimension. In a 2D resistor (in which electrons are said to be confined in a "quantum well") electrons are free to move in two dimensions. In a 3D resistor, electrons are free to move in all three dimensions. According to conventional semiconductor theory (e.g. [1]), we would write the resistances as

$$\begin{aligned} &\text{1D}: \quad R_{1D} = \rho_{1D}\, L \qquad \rho_{1D} = \frac{1}{n_l q \mu_n} \\ &\text{2D}: \quad R_{2D} = \rho_{2D} \frac{L}{W} \qquad \rho_{2D} = \frac{1}{n_s q \mu_n} \\ &\text{3D}: \quad R_{3D} = \rho_{3D} \frac{L}{A} \qquad \rho_{3D} = \frac{1}{n q \mu_n}\,. \end{aligned} \tag{3.1}$$

(Note that the resistivities, ρ, have different units in different dimensions, and the carrier densities, $n_l, n_s,$ and n are per unit length, area, and volume respectively.) These expressions are reasonable; the resistance is proportional to the length of the resistor in each case. It is inversely proportional to the width in 2D or cross-sectional area in 3D because increasing W or A is like adding resistors in parallel. We shall see however, that these equations are not always correct — even for such a simple device, interesting things can happen.

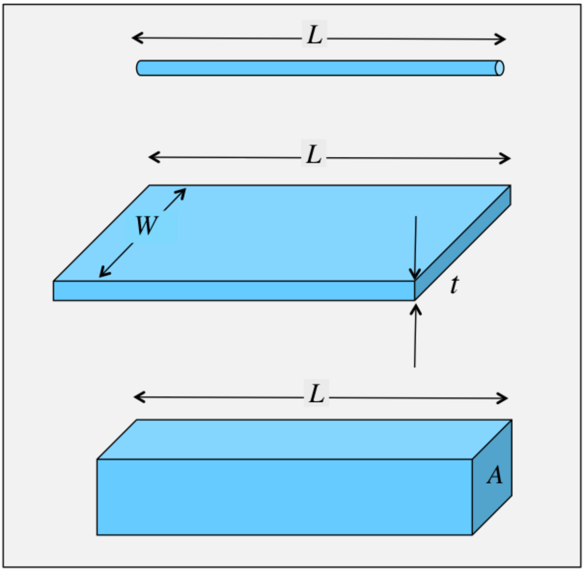

Fig. 3.1. Sketch of 1D, 2D, and 3D resistors. In this chapter, we will focus on 2D resistors, but the same techniques apply in 1D and 3D as well.

Our starting point is the Landauer expression for the conductance, eqn. (2.51), which is repeated below

$$G = \frac{2q^2}{h} \int T(E)M(E)\left(-\frac{\partial f_0}{\partial E}\right) dE \qquad \mathrm{S} = (1/\Omega)\,. \tag{3.2}$$

Equation (3.2) is valid in 1D, 2D, and 3D, as long as we use the correct expression for the number of conducting channels, $M(E)$. To be specific, consider 2D, for which it is convenient to write

$$G = \frac{1}{\rho_{2D}} \frac{W}{L} = \sigma_s \frac{W}{L}\,. \tag{3.3}$$

We will see that for wide and long diffusive conductors, the sheet conductance, σ_s, is independent of W and L. For short conductors, σ_s becomes a function of L and for narrow conductors, the conductance increases with W in a stepwise manner.

In this lecture, we will focus on 2D resistors, just to make the discussion concrete, but similar considerations apply to 1D and to 3D resistors. Recall also that the $-(\partial f_0/\partial E)$ term in eqn. (3.2) came from a Taylor series expansion of $(f_1 - f_2)$ assuming that the temperature of the two contacts was the same. As we will see in the next lecture, there are two driving "forces" for current flow, differences in the Fermi levels of the two contacts (caused by different voltages) and differences in the temperatures of the two contacts. In this lecture we assume that the two contacts are at the same temperature.

3.2 2D resistors: ballistic

We begin with a ballistic resistor, for which $T(E) = 1$. The term, $M(E) = WM_{2D}(E)$ was given in eqn. (2.25) for general bands and by (2.31) for parabolic energy bands, so we just need to understand the term, $-(\partial f_0/\partial E)$, in eqn. (3.2). We refer to this term as the "Fermi window".

Figure 3.2 is a sketch of $f_0(E)$ and $-(\partial f_0/\partial E)$ vs. E. We see that $(-\partial f_0/\partial E)$ is significant only for an energy range of a few $k_B T_L$ near the Fermi level. It is readily shown that the area under the $-(\partial f_0/\partial E)$ vs. E curve is one, so for low temperatures, we may write

$$-\frac{\partial f_0}{\partial E} \approx \delta\left(E - E_F\right) . \tag{3.4}$$

Finally, using eqn. (3.4) with $T(E) = 1$ in eqn. (3.2), we find

$$\boxed{G^{\text{ball}} = \frac{2q^2}{h} M\left(E_F\right) ,} \tag{3.5}$$

which is a general expression valid in any dimension. If the number of channels is small, then we can simply count them, and we find that the conductance or resistance cannot be any value, it is quantized according to

$$R^{\text{ball}} = \frac{1}{M(E_F)} \frac{h}{2q^2} = \frac{12.9 \text{ k}\Omega}{M(E_F)} . \tag{3.6}$$

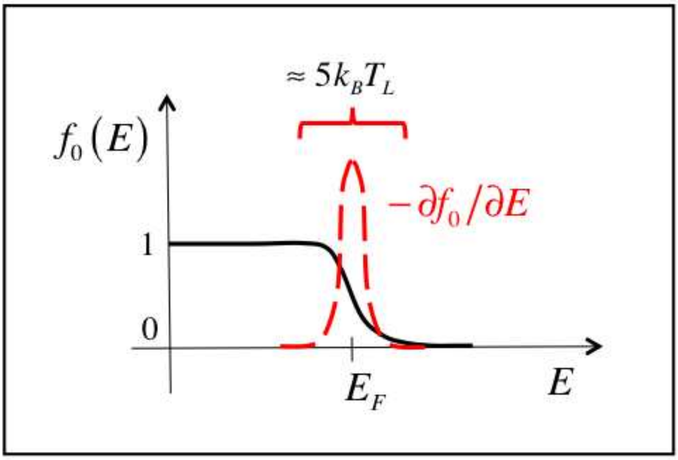

Fig. 3.2. Sketch of the Fermi function and its derivative vs. energy. The function, $-(\partial f_0/\partial E)$, is called the "Fermi window" for conduction.

Note that the ballistic resistance is independent of length, as expected for ballistic transport.

The fact that resistance is quantized is well-established experimentally. See, for example, Fig. 3.3, which shows experimental results. The resistor is a 2D electron gas formed at the interface of AlGaAs and GaAs. The width of the resistor is controlled electrostatically by reverse-biased Schottky junctions. The mobility of the electrons is very high (because the electrons reside in an undoped GaAs layer and because the temperature is low), so ballistic transport is expected. As the width was electrically varied, the measured conductance was seen to increase in discrete steps according to eqn. (3.5). Quantized conductance has been observed in many different systems. The experiment shown in Fig. 3.3 was done at low temperature to achieve near ballistic transport, but modern devices are so short that these effects are becoming important at room temperature in some systems.

Wide, 2D ballistic resistors: $T_L = 0$ K

When W is many electron half-wavelengths, then the number of channels is large, and it is no longer easy to count them. In this limit, we have for

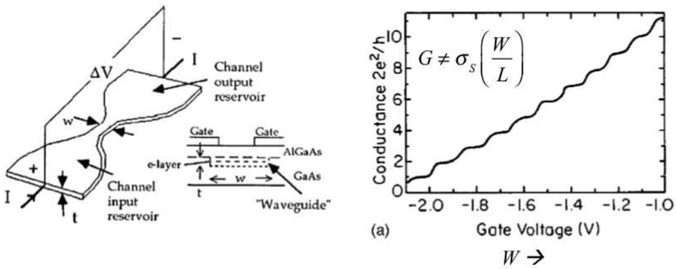

Fig. 3.3. Experiments of van Wees, *et al.* experimentally demonstrating that conductance is quantized. Left: sketch of the device structure. Right: measured conductance. (Data from: B. J. van Wees, *et al.*, *Phys. Rev. Lett.* **60**, 848851, 1988. Figures from D. F. Holcomb, "Quantum electrical transport in samples of limited dimensions", *Am. J. Phys.*, **67**, pp. 278-297, 1999. Reprinted with permission from *Am. J. Phys.* Copyright 1999, American Association of Physics Teachers.)

parabolic energy bands from eqn. (2.31)

$$M(E_F) = WM_{2D}(E_F) = Wg_v \frac{\sqrt{2m^*(E_F - E_c)}}{\pi\hbar} \,. \tag{3.7}$$

It is convenient to relate M_{2D} to the sheet carrier density, n_s because n_s is often known in experiments. In momentum space, all states with $k < k_F$ are occupied at $T_L = 0$. The 2D area of k-space occupied is πk_F^2, each state takes up an area in k-space of $(2\pi)^2/A$, and there is a spin degeneracy of 2, so

$$n_s = g_v \frac{\pi k_F^2}{(2\pi)^2} \times 2 = g_v \frac{k_F^2}{2\pi} \,, \tag{3.8}$$

where g_v is the valley degeneracy. By solving this equation for k_F and using eqn. (2.28), we find

$$M_{2D}(E_F) = \sqrt{\frac{2n_s}{g_v \pi}} \,. \tag{3.9}$$

Equation (3.9) relates the number of channels at the Fermi energy to the sheet carrier density. It is interesting to note that this result does not assume a particular band structure — only that the band is isotropic. To relate k_F to E_F, however, we have to assume a band structure. For example, for parabolic energy bands, the wavevector, k_F is found by solving

$$\frac{\hbar^2 k_F^2}{2m^*} = (E_F - E_c) \,. \tag{3.10}$$

Wide, 2D ballistic resistors: $T_L > 0$ K

At low temperatures, the approximation, eqn. (3.4), works well, but near room temperature and above, we must usually work out the integral in eqn. (3.2). Using the definition of the Fermi function, eqn. (2.1), we find

$$G_{2D}^{\text{ball}} = \frac{2q^2}{h} \int W M_{2D}(E) \left(-\frac{\partial}{\partial E}\right) \frac{1}{1 + e^{(E-E_F)/k_B T_L}} dE \,. \tag{3.11}$$

Integrals of this type appear frequently in semiconductor physics, so let's work this one out as an example.

From the form of the Fermi function, we see that

$$\left(-\frac{\partial}{\partial E}\right) = \left(+\frac{\partial}{\partial E_F}\right) , \tag{3.12}$$

which allows us to move the derivative outside the integral in eqn. (3.11) to obtain

$$G_{2D}^{\text{ball}} = \frac{2q^2}{h} \frac{W g_v \sqrt{2m^*}}{\pi\hbar} \left(\frac{\partial}{\partial E_F}\right) \int_0^{\infty} \frac{\sqrt{(E - E_c)}}{1 + e^{(E-E_F)/k_B T_L}} dE \,, \tag{3.13}$$

where we have used eqn. (3.7) for $M_{2D}(E)$. Next, we change variables by defining

$$\begin{aligned} \eta \;\; &\equiv (E - E_c)/k_B T_L \\ \eta_F &\equiv (E_F - E_c)/k_B T_L \,, \end{aligned} \tag{3.14}$$

and find

$$G_{2D}^{\text{ball}} = \frac{2q^2}{h} \frac{W g_v \sqrt{2m^* k_B T_L}}{\pi\hbar} \left(\frac{\partial}{\partial \eta_F}\right) \int_0^{\infty} \frac{\sqrt{\eta}}{1 + e^{\eta-\eta_F}} d\eta \,. \tag{3.15}$$

The integral in eqn. (3.15) cannot be done analytically, but integrals of this type occur so often in semiconductor work that they have been given a name — Fermi-Dirac integrals. In this case, the integral in eqn. (3.15) is proportional to

$$\mathcal{F}_{1/2}(\eta_F) \equiv \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{\eta^{1/2}}{1 + e^{\eta-\eta_F}} d\eta \,, \tag{3.16}$$

which is the Fermi-Dirac integral of order one-half. Finally, one property of a Fermi-Dirac integral is that when differentiated with respect to its argument, the order of a Fermi-Dirac integral is reduced by one. Putting this all together, we find

$$G_{2D}^{\text{ball}} = \frac{2q^2}{h} \frac{W g_v \sqrt{2m^* k_B T_L}}{\pi\hbar} \left(\frac{\sqrt{\pi}}{2}\right) \mathcal{F}_{-1/2}(\eta_F) = \frac{2q^2}{h} \langle W M_{2D} \rangle \;, \tag{3.17}$$

where

$$\langle M \rangle = \langle W M_{2D} \rangle = \left(\frac{\sqrt{\pi}}{2} \right) W M_{2D}(k_B T_L) \mathcal{F}_{-1/2}(\eta_F)\,, \tag{3.18}$$

where $WM_{2D}(k_BT_L)$ is $WM_{2D}(E-E_c)$ evaluated at an energy of $E-E_c = k_BT_L$. Comparing eqns. (3.17) and (3.5), we see that the conductance at finite temperatures has the same form as the $T_L = 0$ K result — we just replace $M(E_F)$ by $\langle M \rangle$ as defined in eqn. (3.18). We interpret $\langle M \rangle$ as the number of channels in the Fermi window, $(-\partial f_0/\partial E)$.

When analyzing experiments, it is often easier to determine n_s than E_F, but the two are related, so given n_s, we can find E_F. For parabolic bands, the relation is

$$n_s = \int_0^{\infty} D_{2D}(E) f_0(E) dE = g_v \frac{m^* k_B T_L}{\pi \hbar^2} \mathcal{F}_0(\eta_F) = N_{2D} \mathcal{F}_0(\eta_F)\,. \tag{3.19}$$

We have worked out one example assuming a 2D resistor with parabolic energy bands. Similar integrals with Fermi-Dirac integrals of various orders appear when working out problems in other dimensions and for other (i.e. nonparabolic) dispersions. Familiarity with a few properties of Fermi-Dirac integrals is helpful when working out such integrals.

Fermi-Dirac integrals

The Fermi-Dirac integral of order j is defined as

$$\mathcal{F}_j(\eta_F) \equiv \frac{1}{\Gamma(j+1)} \int_0^{\infty} \frac{\eta^j \, d\eta}{1 + e^{\eta - \eta_F}}\,, \tag{3.20}$$

where the Γ-function is defined for integer arguments of zero or greater as

$$\Gamma(n) = (n-1)!\,. \tag{3.21}$$

We also have the following useful relations,

$$\begin{aligned} \Gamma(1/2) &= \sqrt{\pi} \\ \Gamma(p+1) &= p\Gamma(p) \end{aligned} \tag{3.22}$$

For non-degenerate semiconductors, the Fermi level is several k_BT_L below the band edge, so $\eta_F = (E_F - E_c)/k_BT_L \ll 0$. Under these conditions, Fermi-Dirac integrals of any order reduce to exponentials:

$$\mathcal{F}_j(\eta_F) \to e^{\eta_F} \qquad \eta_F \ll 0\,. \tag{3.23}$$

Another useful property that we have already seen involves taking the derivative of a Fermi-Dirac integral,

$$\frac{d\mathcal{F}_j(\eta_F)}{d\eta_F} = \mathcal{F}_{j-1}(\eta_F)\,. \tag{3.24}$$

These few definitions and rules are all we need for most semiconductor problems. One warning — don't confuse the "script" Fermi-Dirac integral as defined in eqn. (3.20) with the "Roman" Fermi-Dirac integral, $F_j(\eta_F)$, which does not include the Γ-function normalization. For a good introduction to Fermi-Dirac Integrals — including approximations and scripts to compute them, see [2].

Exercise 3.1: Analysis of a silicon MOSFET

Let's see how we can use the results of this section to analyze the performance of modern field-effect transistors. Figure 3.4 shows the measured I–V characteristics of a 60 nm channel length silicon MOSFET. Let's focus on the low drain voltage (near-equilibrium) region and the $V_G = 1.2$ V characteristic (the top line). For this condition, the measured carrier density and channel resistance (after subtracting out the parasitic series resistances of the source and drain contacts) are

$$n_s \approx 6.7 \times 10^{12}\ \text{cm}^{-2}$$

$$R_{ch} \approx 215\ \Omega\text{-}\mu\text{m}$$

$$\mu_{\text{eff}} \approx 260\ \text{cm}^2/\text{V-s}\,.$$

The two questions we ask are: 1) How many conduction channels carry the current? and 2) How close is the channel resistance to the ballistic limit? Before we can answer these questions, we need to understand a little bit about the Si band structure. Recall that in bulk Si, there are six equivalent conduction band minima [1], but quantum confinement lifts the degeneracy so that the lowest subband is two-fold degenerate ($g_v = 2$) with an effective mass of $m^* = m_t = 0.19m_0$ [3]. We can do the calculation in three different ways. First, we can assume $T_L = 0$ K, which makes the math easy, but the $T_L = 0$ K assumption is not so good at room temperature. Second, we can assume Maxwell-Boltzmann (non-degenerate) carrier statistics, which also results in simple math, but above the threshold voltage, the non-degenerate assumption is not so good. The third way to do the calculation is without either assumption, which entails evaluating Fermi-Dirac integrals. In this

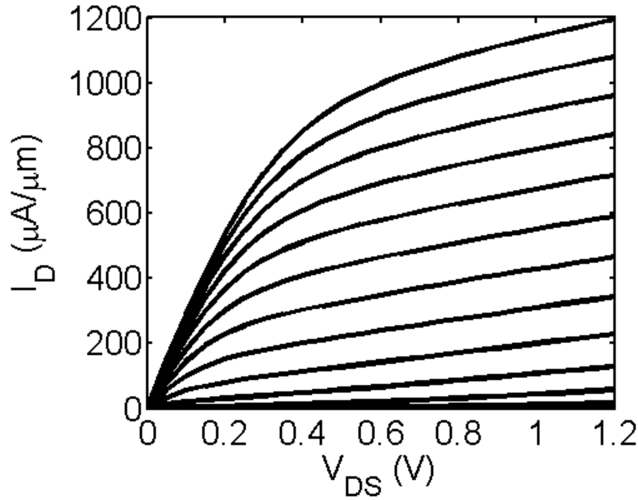

Fig. 3.4. Measured $I_D - V_{DS}$ characteristic of an n-channel silicon MOSFET. (Data from: Changwook Jeong, Dimitri A. Antoniadis and Mark Lundstrom, "On Back-Scattering and Mobility in Nanoscale Silicon MOSFETs", *IEEE Trans. Electron Dev.*, **56**, 2762-2769, 2009.)

simple example, we shall consider just the first approach, which should give us a good feel for the numbers.

Assuming $T_L = 0$ K, eqn. (3.9), gives $M_{2D}(E_F) \approx 150/\mu$m. Consider a minimum size transistor with $W/L = 2$. Since $L = 60$ nm, we find $M_{2D}(E_F) \approx 18$. A rather small number of channels carry the current in a minimum sized MOSFET. To find the corresponding ballistic resistance of this MOSFET, we use eqn. (3.6) to find $R_{2D}^{\text{ball}} \approx 90\ \Omega$-$\mu$m, which is almost one-half of the channel resistance of this MOSFET.

3.3 2D resistors: diffusive to ballistic

In the previous section, we assumed ballistic transport, $T(E) = 1$. In this section, we will first assume diffusive transport, $T(E) = \lambda(E)/L$. Our general approach works for 1D, 2D, and 3D resistors, but as in the previous section, we'll just work out the 2D case here. Our starting point is eqn. (3.2) in 2D and in the diffusive limit.

$$G_{2D}^{\text{diff}} = \left(\frac{2q^2}{h} \int \lambda(E) M_{2D}(E) \left(-\frac{\partial f_0}{\partial E} \right) dE \right) \frac{W}{L} \qquad (1/\text{Ohm})\,. \qquad (3.25)$$

The ratio, W/L, is consistent with conventional transport theory (e.g. eqn. (3.1)). In our Landauer picture, the W enters because the number of channels is proportional to W, and the L comes because in the diffusive limit, the transmission is proportional to $1/L$. Since the integral in eqn. (3.25) is trivial for $T_L = 0$ K, we begin there.

Wide, 2D diffusive resistors: $T_L = 0$ K

Using eqn. (3.4), we find that for $T_L = 0$ K, eqn. (3.25) becomes

$$G_{2D}^{\text{diff}} = \frac{2q^2}{h} M_{2D}\left(E_F\right) \lambda\left(E_F\right) \frac{W}{L} = \frac{\lambda(E_F)}{L} G_{2D}^{\text{ball}}\,, \tag{3.26}$$

so if we know the mean-free-path, it is easy to compute the diffusive conductance.

We have discussed the ballistic and diffusive conductances, but it is also easy to cover the entire ballistic to diffusive spectrum. By using $T(E) = \lambda(E)/(\lambda(E) + L)$, we find

$$G_{2D} = \frac{2q^2}{h} W M_{2D}\left(E_F\right) \frac{\lambda(E_F)}{\lambda(E_F) + L} = \frac{\lambda(E_F)}{\lambda(E_F) + L} G_{2D}^{\text{ball}}\,, \tag{3.27}$$

or in terms of resistance,

$$R = \left(1 + \frac{L}{\lambda(E_F)}\right) R^{\text{ball}}\,. \tag{3.28}$$

Equation (3.28) shows that $R \propto L$ in the diffusive limit and R is independent of L in the ballistic limit.

Wide, 2D diffusive resistors: $T_L > 0$ K

At finite temperatures, we should work out the integral in eqn. (3.25). By multiplying and dividing eqn. (3.25) by

$$\langle M_{2D} \rangle \equiv \int M_{2D}(E) \left(-\frac{\partial f_0}{\partial E}\right) dE\,,$$

we can re-arrange it as,

$$G_{2D}^{\text{diff}} = \frac{2q^2}{h} \langle M_{2D} \rangle \langle\langle \lambda \rangle\rangle \frac{W}{L} = \frac{\langle\langle \lambda \rangle\rangle}{L} G_{\text{ball}}^{2D}\,, \tag{3.29}$$

where $\langle M \rangle$ was defined in eqn. (3.18), and the average mean-free-path is defined as

$$\langle\langle \lambda \rangle\rangle \equiv \frac{\int \lambda(E) M_{2D}(E) \left(-\frac{\partial f_0}{\partial E}\right) dE}{\int M_{2D}(E) \left(-\frac{\partial f_0}{\partial E}\right) dE} = \frac{\langle M\lambda \rangle}{\langle M \rangle} . \qquad (3.30)$$

Equation (3.29) has the same form as (3.26). For $T_L = 0$ K, we just replace $\langle M \rangle$ with $M(E_F)$ and $\langle\langle \lambda \rangle\rangle$ with $\lambda(E_F)$. The single and double brackets are just to remind us that we are dealing with two different, specially defined averages.

To actually evaluate the average mean-free-path, we need to assume a band structure and specify $\lambda(E)$. A simple way to write $\lambda(E)$ for some common scattering mechanisms is in so-called power-law form [4],

$$\lambda(E) = \lambda_0 \left(\frac{E - E_c}{k_B T_L}\right)^r , \qquad (3.31)$$

where r is a characteristic exponent describing a particular scattering process and λ_0 is a (typically temperature-dependent) constant. For example, in 3D, $r = 0$ for carrier scattering by acoustic phonons, and $r = 2$ for ionized impurity scattering.

With only a little straight-forward (and not too tedious) math, we can work out the integral in eqn. (3.30) to find

$$\langle\langle \lambda \rangle\rangle = \lambda_0 \times \left(\frac{\Gamma\,(r + 3/2)}{\Gamma\,(3/2)}\right) \times \left(\frac{\mathcal{F}_{r-1/2}(\eta_F)}{\mathcal{F}_{-1/2}(\eta_F)}\right) . \qquad (3.32)$$

If $r = 0$ the mean-free-path is independent of energy and $\langle\langle \lambda \rangle\rangle = \lambda_0$.

Finally, we should consider the entire ballistic to diffusive spectrum. For an energy dependent mean-free-path, the result is a bit complicated, but for a constant mean-free-path, the result is much like eqn. (3.28)

$$R = R^{\text{ball}} \left(1 + \frac{L}{\lambda_0}\right) . \qquad (3.33)$$

Exercise 3.2: Analysis of a silicon MOSFET (continued)

In the previous example, we estimated the ballistic channel resistance by assuming $T_L = 0$ K (not because it is a good assumption, just to keep the math simple) and found $R^{\text{ball}} \approx 90\ \Omega$-$\mu$m. The measured channel resistance is 215 Ω-μm, so from eqn. (3.33) we have

$$215 = \left(1 + \frac{L}{\lambda_0}\right) 90 \rightarrow \lambda_0 \approx 40\,\text{nm}$$

Our $T_L = 0$ K assumption for analyzing room temperature data results in a mean-free-path that is a little too large (assuming Maxwell Boltzmann statistics gives 15 nm), but this simple calculation does illustrate the point that modern silicon transistors are in the quasi-ballistic transport regime — neither fully ballistic nor fully diffusive. A more careful analysis would involve Fermi-Dirac integrals and also consider the possibility that multiple subbands are occupied.

3.4 Discussion

In this lecture we discussed the evaluation of the resistance of a 2D conductor. The techniques and concepts apply to other dimensions as well. In this section, we will discuss a few topics relating to resistors and resistivity.

A few words about mobility

The conventional description of resistance begins with eqns. (3.1), but these equations do not apply to ballistic or quasi-ballistic resistors, and it is not always clear about how to evaluate the mobility. Our approach to transport begins with eqn. (3.2), which applies from the ballistic to diffusive limits. There has been no need for us to discuss mobility. Although mobility is a commonly-used concept, it can often be confusing. For example, eqns. (3.1) tell us that the conductivity is proportional to the electron density times the mobility, but the term, $-(\partial f_0/\partial E)$, in eqn. (3.2) (the Fermi window) ensures that only electrons near the Fermi level contribute to current flow. For an n-type semiconductor, this is sometimes all of the carriers in the conduction band (non-degenerate semiconductors), but sometimes only a small fraction of them (degenerate semiconductors). Nevertheless, because the mobility concept is so frequently used, we should discuss it.

The best way to define the mobility is to begin with eqn. (3.2) and equate it to the conventional expression,

$$G_{2D} = \frac{2q^2}{h} \int T(E)M(E) \left(-\frac{\partial f_0}{\partial E}\right) dE \equiv n_s q \mu \frac{W}{L}\,, \qquad (3.34)$$

from which we find the mobility as

$$\mu_{\text{app}} \equiv \frac{1}{n_s}\frac{2q}{h}\int T(E)L\,M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE\,, \tag{3.35}$$

which we take as the definition of mobility, not the simple Drude expression,

$$\mu = \frac{q\tau}{m^*}\,, \tag{3.36}$$

where τ is the momentum relaxation time. We label the mobility in eqn. (3.35) an "apparent" mobility because eqn. (3.35) is defined for ballistic transport as well as diffusive. For example, setting $T(E) = 1$, we find the "ballistic mobility" as

$$\mu_{\text{ball}} = \frac{1}{n_s}\frac{2q}{h}\int LM_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE\,. \tag{3.37}$$

Similarly, setting $T(E) = \lambda(E)/L$, we find the traditional, diffusive mobility as

$$\boxed{\mu_{\text{diff}} = \frac{1}{n_s}\frac{2q}{h}\int \lambda(E)M_{2D}(E)\left(-\frac{\partial f_0}{\partial E}\right)dE\,.} \tag{3.38}$$

The concept of a ballistic mobility was introduced by Shur [5] and can be useful in device analysis. It allows us to use traditional expressions in the ballistic limit simply by replacing the actual mobility with the ballistic mobility. Comparing eqn. (3.37) with (3.38), we see that the ballistic mobility is the diffusive mobility with the mean-free-path replaced by the length of the resistor. Recall that in the contacts, strong scattering maintains equilibrium. An electron injected into a ballistic channel last scattered in the first contact, and then scatters next in the second contact. The distance between scattering events is, therefore, the channel length, so it makes physical sense that it plays the role of the mean-free-path in the ballistic mobility.

Increasingly in nanoscale electronics, problems lie between the ballistic and diffusive limits. In this case, $T(E) = \lambda(E)/(\lambda(E) + L)$, and we can show that the apparent mobility is

$$\frac{1}{\mu_{\text{app}}} = \frac{1}{\mu_{\text{diff}}} + \frac{1}{\mu_{\text{ball}}}\,, \tag{3.39}$$

which looks like a traditional Mathiessen's Rule for combining mobilities [4]. Another way to look at this is to use eqn. (3.38) for the diffusive mobility but replace the mean-free-path with an apparent mean-free-path,

$$\frac{1}{\lambda_{\text{app}}} = \frac{1}{\lambda} + \frac{1}{L} \,. \tag{3.40}$$

The apparent mean-free-path is the actual mean-free-path due to scattering or the length of the channel — whichever is shorter.

Assuming parabolic energy bands, the mobilities for general conditions can be worked out in terms of Fermi-Dirac integrals. We leave that as an exercise. Consider the simpler case of $T_L = 0$ K. The electron density in the conduction band is

$$n_s = g_v \frac{m^*}{\pi\hbar^2}\left(E_F - E_c\right) = D_{2D}\left(E_F - E_c\right) \,, \tag{3.41}$$

and we can also show

$$M_{2D} = \frac{h}{4}\left\langle \upsilon_x^+ \right\rangle D_{2D} \,, \tag{3.42}$$

where

$$\left\langle \upsilon_x^+ \right\rangle = \frac{2}{\pi}\upsilon_F \,, \tag{3.43}$$

with υ_F being the Fermi velocity.

Using eqns. (3.37), (3.38), (3.41), and (3.42), we find

$$\begin{aligned} \mu_{\text{diff}} &= \frac{D_{\text{diff}}}{\left(E_F - E_c\right)/q} \\ \mu_{\text{ball}} &= \frac{D_{\text{ball}}}{\left(E_F - E_c\right)/q} \,, \end{aligned} \tag{3.44}$$

where the diffusion coefficients are given by

$$\begin{aligned} D_{\text{diff}} &= \frac{\langle \upsilon_x^+ \rangle \lambda(E_F)}{2} \\ D_{\text{ball}} &= \frac{\langle \upsilon_x^+ \rangle L}{2} \,. \end{aligned} \tag{3.45}$$

Equations (3.44) look like Einstein relations between the mobility and diffusion coefficient with $(E_F - E_c)/q$ playing the role of $k_B T_L/q$ since we have assumed $T_L = 0$ K. Note that in eqn. (3.45), D_{diff} is what is generally called the diffusion coefficient, and we have also defined a "ballistic diffusion coefficient"!

Exercise 3.3: Analysis of a silicon MOSFET (continued)

Let's return to our MOSFET example and estimate the ballistic mobility of this transistor. The measured mobility comes from a long channel MOSFET, so it is the traditional, diffusive mobility. Let's estimate the ballistic mobility. We'll assume $T_L = 0$ K again, to keep the math simple. Under this assumption, the ballistic mobility, eqn. (3.44) can be expressed in terms of the inversion layer density, n_s according to

$$\mu_{\text{ball}} = \frac{2q}{h} L \sqrt{2g_v/\pi n_s} \,. \tag{3.46}$$

Inserting numbers, we find, $\mu_{\text{ball}} \approx 1200$ cm^2/V-s, which is larger than the diffusive mobility. According to eqn. (3.39), the apparent mobility of this MOSFET will be a little lower than the bulk mobilty because of the ballistic mobility. (The precise numbers here should be taken with caution. We are assuming $T_L = 0$ K to keep the math simple but the mobility was measured at 300 K.)

Ways to write the conductivity

According to eqn. (3.26), the 2D diffusive conductivity (also called the sheet conductance) at $T_L = 0$ K is

$$\sigma_s = \frac{2q^2}{h} M_{2D}\left(E_F\right) \lambda\left(E_F\right) \,. \tag{3.47}$$

You will see this expression written in several different ways, so it's worth mentioning some of the common forms.

In Lecture 2, we saw that

$$\begin{aligned} M_{2D}(E) &= \frac{h}{4} \left\langle v_x^+ \right\rangle D_{2D}(E) \\ \left\langle v_x^+ \right\rangle &= \frac{2}{\pi} v \,, \end{aligned} \tag{3.48}$$

and in Lecture 6 we will learn that

$$\lambda(E) = \frac{\pi}{2} v(E) \tau_m(E) \,, \tag{3.49}$$

where τ_m is the momentum relaxation time, the time between electron scattering events. Using these expressions, we can rewrite eqn. (3.47) as

$$\sigma_s = q^2 D_{2D}\left(E_F\right) \frac{v^2(E_F)\tau(E_F)}{2} \,, \tag{3.50}$$

which is a form that you will commonly see. By defining the electron diffusion coefficient as

$$D_n = \frac{\upsilon^2(E_F)\tau(E_F)}{2}\,, \tag{3.51}$$

we can rewrite eqn. (3.50) as

$$\sigma_s = q^2 D_{2D}\left(E_F\right) D_n(E_F)\,, \tag{3.52}$$

which is also a common way to write the sheet conductance.

Finally, let's discuss one more way. Assuming parabolic energy bands, we have

$$\frac{1}{2}m^*\upsilon(E_F)^2 = (E_F - E_c)\ . \tag{3.53}$$

Using this and eqn. (3.41) for n_s, we can write eqn. (3.50) as

$$\sigma_s = n_s q \mu_n\ , \tag{3.54}$$

where the mobility is

$$\mu_n = \frac{q\tau(E_F)}{m^*}\,. \tag{3.55}$$

Equation (3.55) is a familiar result, but not a good starting point for analysis in general. For example, in research, we frequently encounter problems for which the parabolic band assumption that leads to an effective mass, m^* is not appropriate.

We summarize this discussion by collecting the various ways to write the 2D conductivity:

$$\boxed{\begin{aligned}
\sigma_s &= \frac{2q^2}{h}M_{2D}\left(E_F\right)\lambda\left(E_F\right)\\
\sigma_s &= q^2 D_{2D}\left(E_F\right)\frac{\upsilon^2(E_F)\tau(E_F)}{2}\\
\sigma_s &= q^2 D_{2D}\left(E_F\right) D_n(E_F)\\
\sigma_s &= n_s q\mu_n = n_s q\left(\frac{q\tau(E_F)}{m^*}\right)\ .
\end{aligned}} \tag{3.56}$$

While these expressions are all equivalent, they provide different perspectives. According to the second and third forms, what's important is the density of states and velocity or diffusion coefficient at the Fermi level. According to the last expression, the conductivity is proportional to the total carrier density. Of course the term, $-(\partial f_0/\partial E)$, in the conductivity expression tells us that the current is carried by electrons near the Fermi level (i.e. inside the Fermi window), but the position of the Fermi level can be related to the total carrier density. It is useful to understand these different perspectives because sometimes a problem that is difficult from one perspective is easy when viewed from a different perspective.

Finally, we should extend our discussion to finite temperatures where eqn. (3.47) becomes

$$\sigma_s = \frac{2q^2}{h} \int M_{2D}(E)\,\lambda(E) \left(-\frac{\partial f_0}{\partial E}\right) dE\,. \tag{3.57}$$

It is convenient to define a "differential conductivity" so that this equation can be written as

$$\boxed{\begin{aligned} \sigma_s &= \int \sigma_s'(E)\, dE \\ \sigma_s' &= \frac{2q^2}{h} M_{2D}(E)\,\lambda(E) \left(-\frac{\partial f_0}{\partial E}\right). \end{aligned}} \tag{3.58}$$

To find the total conductivity, we add up the contributions from each energy channel. You will find eqn. (3.58) written in several different ways, as in eqn. (3.56).

Where is the power dissipated?

For any resistor, the power dissipated is $P_d = VI = V^2/R$. The power input to the electron system from the battery is typically dissipated by electron-phonon scattering, which transfers the energy to the lattice and heats it up. For a ballistic resistor, there is no scattering in the channel, but the power dissipated is still V^2/R. Where is this power dissipated? Not surprisingly, if the answer isn't in the channel, then it must be in the contacts.

We know that current flows in the Fermi window, the energy range where $f_1 - f_2$ is non zero. At $T_L = 0$ K, the current flows between the two Fermi levels, as illustrated in Fig. 3.5. For finite temperatures and small applied biases, the current flows in the region where $-(\partial f_0/\partial E)$ is significant. As shown in Fig. 3.5, when an electron leaves contact 1, it leaves a hole (an empty state) in the contact. These electrons enter contact 2 with some excess energy (we say that they are "hot" electrons), which they lose by

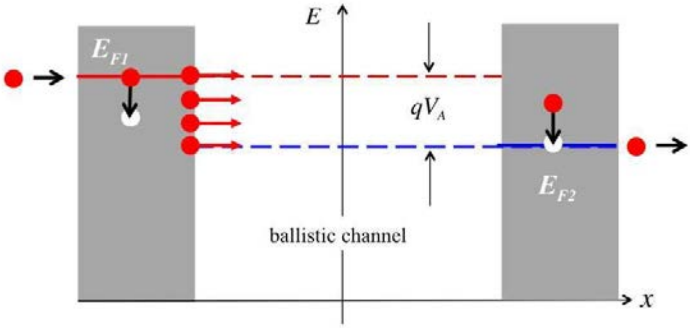

Fig. 3.5. Illustration of how power is dissipated in a ballistic resistor.

inelastic scattering in contact 2. The average energy of the electrons that enter contact 2 is about $qV_A/2$, so about half of the power is dissipated in contact 2. Current flows near the Fermi level, so charge neutrality in contact 2 is maintained when an electron at the Fermi energy leaves contact 2 and travels around the external circuit until it enters contact 1. The electron enters contact 1 at the Fermi energy, and inelastic scattering processes fill up the hole that was created when the first electron was injected into the channel. On average, the energy of the hole is about $qV_A/2$, so the other half of the power is dissipated in contact 1.

We conclude that in a ballistic channel near-equilibrium, one-half of the power is dissipated in each of the two contacts.

Where does the voltage drop?

In a diffusive resistor, the voltage drops linearly across the length of the resistor. Where does the voltage drop in a ballistic resistor? Not surprisingly, the answer is at the contacts. For a thorough discussion of this topic, see Chapter 2 in Datta [6]. Here, we just explain why it is reasonable to expect the voltage to drop across the two contacts.

Recall that a voltmeter measures differences in the Fermi levels (electrochemical potentials) of the two contacts. For example, in a p-n junction, there is an electrostatic potential difference between the p and n-type regions in equilibrium, but a voltmeter reads zero because the Fermi level is the same in the two contacts.

Figure 3.6 is a sketch of an energy band diagram for a ballistic resistor under bias. $E_{F2} = E_{F1} - qV_A$, so a voltmeter across the two contacts

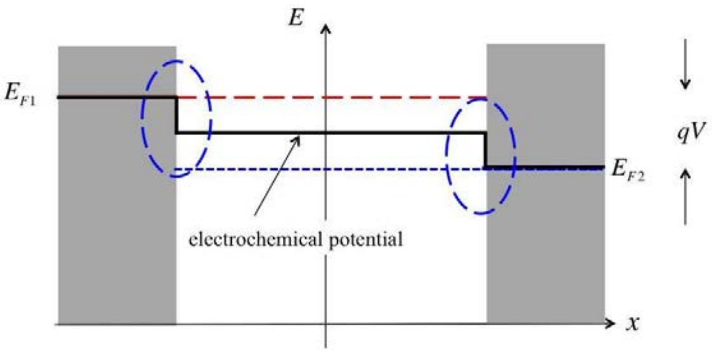

Fig. 3.6. Energy band diagram of a ballistic resistor under bias illustrating how we associate the internal voltage drop with the change in the electrochemical potential (also known as the quasi-Fermi level).

will register a voltage of V_A, but where does this voltage drop within the resistor?

Inside contact 1, there is one, well-defined Fermi level, E_{F1}, and inside contact 2, there is a second, well-defined Fermi level, E_{F2}. Inside the device however, there are two Fermi levels. Some states are filled by the source. Since they are in equilibrium with the source, they are filled according to a Fermi function with the source Fermi level. The other set of states is filled by the drain according to a Fermi function with the drain Fermi level. If we compute the average Fermi level, the electrochemical potential, it has the shape shown by the solid line in Fig. 3.6. We see that one-half of the electrochemical potential drop occurs at the first contact and the other half at the second contact. Since a voltmeter responds to changes in the electrochemical potential, we conclude that the voltage drops equally at the two contacts. For this reason, the ballistic resistance, 12.9 k$\Omega/\langle M \rangle$, is often called the "quantum contact resistance".

Resistors in 1D and 3D

Our general expression for conductance, eqn. (3.2), works for the 1D, 2D, or 3D resistors sketched in Fig. 3.1, as long as we properly count the modes (channels) for current flow. Consider again the 2D resistor sketched in Fig. 3.1. It is long in the direction of current flow, but finite in the width and thickness directions. Whenever electrons are confined in a potential well, their energies are quantized into "particle in a box" states according to

$$\epsilon_n = \frac{\hbar^2 \pi^2 n^2}{2m^* a^2}\,, \tag{3.59}$$

where a is the width of the potential well, and n is an integer. Each of these "subbands" is occupied according to its location with respect to the Fermi level and is a channel for current flow. If the thickness of the 2D sheet is very thin, then the corresponding subbands are widely spaced in energy, and we can count them. (We have been implicitly assuming that only the lowest subband is occupied.) If the width, W, of the 2D sheet is large, then the corresponding subbands are closely spaced in energy, and many are occupied. In this case, the number of subbands is proportional to W. To include all subbands, we write

$$M(E) = WM_{2D}\left(E\right) = \sum_{n=1}^{N} W g_v \frac{\sqrt{2m^*\left(E-\epsilon_n\right)}}{\pi\hbar}\,, \tag{3.60}$$

where the sum is over the subband in the vertical confinement direction.

Now consider the 1D resistor in Fig. 3.1; it is like a very narrow 2D resistor. If the width and thickness are both small, then all of the subbands are widely spaced in energy, and we can simple count them,

$$M(E) = M_{1D}\left(E\right) = \text{No. of subbands at energy, } E\,. \tag{3.61}$$

Finally, if both the width and thickness of the resistor are large, all of the subbands are closely spaced in energy, and

$$M(E) = AM_{3D}\left(E\right) = g_v \frac{m^*\left(E-E_c\right)}{2\pi\hbar^2}\,. \tag{3.62}$$

For a 1D (nanowire) resistor, we have strong quantum confinement in two dimensions, and $M(E)$ is given by eqn. (3.61). For a 2D (planar) resistor, we have strong quantum confinement in one dimension, and $M(E)$ is given by eqn. (3.60). And for a 3D resistor, there is no quantum confinement and $M(E)$ is given by eqn. (3.62). (Note that eqns. (3.60) and (3.62) assume parabolic energy bands.) Having defined $M(E)$, we can evaluate eqn. (3.2) in any dimension.

It is convenient to rewrite eqn. (3.2) as

$$\boxed{\begin{aligned} G &= \frac{2q^2}{h}\langle\langle T\rangle\rangle\langle M\rangle \\ \langle M\rangle &= \int M(E)\left(-\frac{\partial f_0}{\partial E}\right)dE \\ \langle\langle T\rangle\rangle &= \frac{\int T(E)M(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}{\int M(E)\left(-\frac{\partial f_0}{\partial E}\right)dE}\,. \end{aligned}} \tag{3.63}$$

If we assume a constant mean-free-path (just to keep things simple) we can write the resistance as

$$G = \frac{2q^2}{h}\frac{\lambda_0}{\lambda_0 + L}\langle M\rangle\ . \tag{3.64}$$

Assuming parabolic energy bands, we can write in 1D:

$$\langle M_{1D}\rangle = \sum_i \mathcal{F}_{-1}\left(\eta_{Fi}\right)\ , \tag{3.65}$$

where

$$\eta_{Fi} = \frac{E_F - \epsilon_i}{k_B T_L}\,, \tag{3.66}$$

and the index, i, refers to the various subbands. At $T_L = 0$ K, the 1D expressions reduce to

$$G_{1D} = \frac{2q^2}{h}\frac{\lambda_0}{\lambda_0 + L} \times \text{No. of subbands at energy, } E_F\ . \tag{3.67}$$

For Maxwell-Boltzmann statistics, we find

$$G_{1D} = n_l q \mu_{\text{app}}\frac{1}{L}\,, \tag{3.68}$$

where

$$\begin{aligned} \mu_{\text{app}} &= \frac{D_n}{(k_B T_L/q)} \\ D_n &= \upsilon_T\lambda_{\text{app}}/2 \\ \upsilon_T &= \sqrt{\left(2k_B T_L\right)/\pi m^*} \\ \frac{1}{\lambda_{\text{app}}} &= \frac{1}{\lambda_0} + \frac{1}{L}\,. \end{aligned} \tag{3.69}$$

In 2D, we find

$$\langle M \rangle = W \langle M_{2D} \rangle = \frac{\sqrt{\pi}}{2} W M_{2D}(k_B T_L) \sum_i \mathcal{F}_{-1/2}(\eta_{Fi}) \ , \tag{3.70}$$

where

$$M_{2D}(k_B T_L) = g_v \frac{\sqrt{2m^* k_B T_L}}{\pi\hbar} \, . \tag{3.71}$$

At $T_L = 0$ K, eqn. (3.65) simplifies to

$$G_{2D} = \frac{2q^2}{h} \frac{\lambda_0}{\lambda_0 + L} W M_{2D}(E_F) \ . \tag{3.72}$$

For Maxwell-Boltzmann statistics, we find

$$G_{2D} = n_s q \mu_{\text{app}} \frac{W}{L} \, . \tag{3.73}$$

Finally, in 3D, the expressions become:

$$\langle M \rangle = A \langle M_{3D} \rangle = A M_{3D}(k_B T_L) \mathcal{F}_0(\eta_F) \ , \tag{3.74}$$

where

$$M_{3D}(k_B T_L) = g_v \frac{m^* k_B T_L}{2\pi\hbar^2} \tag{3.75}$$

and

$$\eta_F = \frac{E_F - E_c}{k_B T_L} \, . \tag{3.76}$$

At $T_L = 0$ K, the 3D expressions reduce to

$$G_{3D} = \frac{2q^2}{h} \frac{\lambda_0}{\lambda_0 + L} A M_{3D}(E_F) \ . \tag{3.77}$$

For Maxwell-Boltzmann statistics, we find

$$G_{3D} = n q \mu_{\text{app}} \frac{A}{L} \, . \tag{3.78}$$

Exercise 3.4: A 1D example

Let's end this chapter with a simple example. Carbon nanotubes are almost ideal 1D conductors. Figure 3.7 is the measured IV characteristic of a metallic carbon nanotube. For low voltages, the current is linear, and we can read off the conductance as 22 μS. For this material, the $T_L = 0$ K approximation is good even at moderate temperatures, so we can use eqn. (3.67). There is a valley degeneracy of two for the carbon nanotube, and assuming that one subband is occupied, we find the ballistic resistance to be $G_B = 154$ μS. From eqn. (3.67), we can estimate the mean-free-path to be $\lambda_0 = 167$ nm, which is much less than the 1 μm length of the carbon nanotube, so transport in this carbon nanotube is diffusive.

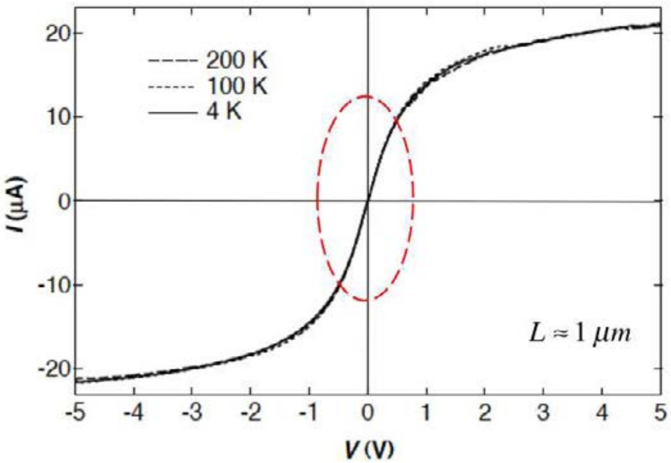

Fig. 3.7. Measured I–V characteristic of a metallic carbon nanotube. Essentially identical measurements at $T_L = 4, 100$, and 200 K are shown. (Zhen Yao, Charles L. Kane, and Cees Dekker, "High-Field Electrical Transport in Single-Wall Carbon Nanotubes", *Phys. Rev. Lett.*, **84**, 2941-2944, 2000. Reprinted with permission from *Phys. Rev. Lett.* Copyright 2000, American Physical Society.)

3.5 Summary

Our goal in this lecture has been to learn how to use eqn. (3.2), which applies quite generally when the temperature is uniform. We learned that:

(1) Conductors display a finite resistance, even in the absence of scattering in the resistor. The ballistic resistance sets a lower limit to resistance no matter how short the resistor is. This ballistic limit is becoming important in practical, room temperature devices.
(2) The ballistic (quantum contact) resistance is quantized, and the quantum of resistance is $h/2q^2$.
(3) Transport from the ballistic to diffusive limit is easily treated by using the transmission.
(4) Resistors in 1D, 2D, or 3D can all be treated with a common formalism.

For the most part, we assumed parabolic energy bands with an effective mass, m^*, when working out results. It's important to realize, however, that the assumption of parabolic energy bands is not necessary. For an example of working with a much different bandstructure, see Lecture 10, where we develop similar expressions for graphene, a material with linear, not parabolic energy bands.

Finally, the key point to remember is that when you encounter a new material or nanostructure and need to know its resistance, the place to begin is at eqn. (3.2) not at eqn. (3.1).

3.6 References

For an introduction to semiconductor theory as used for analyzing electronic devices, see:

[1] Robert F. Pierret *Advanced Semiconductor Fundamentals, 2nd Ed.*, Vol. VI, Modular Series on Solid-State Devices, Prentice Hall, Upper Saddle River, N.J., USA, 2003.

The essentials of Fermi-Dirac integrals are described in:

[2] R. Kim and M.S. Lundstrom, "Notes on Fermi-Dirac Integrals", 3rd Ed., https://www.nanohub.org/resources/5475.

For a discussion of quantum confinement in silicon MOSFETs, how the valley degeneracy is lifted, etc., see:

[3] M. Lundstrom, "ECE 612: Nanoscale Transistors (Fall 2008). Lecture

4: Polysilicon Gates/QM Effects", http://nanohub.org/resources/5364, 2008.

Mathiessen's Rule for combining mobilities is discussed in Chapter 2 of Lundstrom, and power law scattering is disucssed on pp. 67 and 137.

[4] Mark Lundstrom, *Fundamentals of Carrier Transport 2nd Ed.*, Cambridge Univ. Press, Cambridge, UK, 2000.

The concept of ballistic mobility was introduced by Michael Shur in

[5] M.S. Shur, Low Ballistic Mobility in GaAs HEMTs, *IEEE Electron Dev. Lett.*, **23**, 511-513, 2002.

For an excellent discussion of power dissipation and voltage drops in ballistic conductors, see:

[6] Supriyo Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge Univ. Press, Cambridge, UK, 1995.