ELECTRON TRANSPORT AT THE NANOSCALE
Lecture Notes, preliminary version

Geert Brocks

December 2005

Chapter 1

Electron transport in one dimension

This chapter gives a simple introduction to scattering problems in one dimension and their relation to the conductance of one-dimensional systems at the level of introductory quantum mechanics. Purely one-dimensional systems are of limited practical use, but their simplicity allows one to introduce the basic physics using simple mathematics only. Sec. 1.1 recapitulates the concept of a quantum current, as defined in introductory quantum mechanics. The Landauer formula expresses the electronic conductance or resistance of a system as a scattering problem.[1] A naive derivation is given in Sec. 1.2.[2] Sec. 1.3 introduces some elementary folklore of scattering theory, the transfer matrix and the scattering matrix, using the ubiquitous square barrier potential as an example. Techniques to solve scattering problems for any one-dimensional potential are introduced in Sec. 1.4. These will be generalized to three-dimensional problems in the next chapter. In particular I focus on so-called "mode matching", which is the most accessible technique. I discuss its relation to the "Green function" and the "embedding" techniques, two other currently used methods. The techniques are illustrated by a simple model, which consists of a straightforward discretization of the one-dimensional Hamiltonian.[3] This model is mathematically equivalent to a simple tight-binding model of a quantum wire of atoms, as is shown in Sec. 1.5.[4]

1.1 Quantum currents

1.1.1 Probability currents

The **probability current** $J(x, t)$

$$J(x, t) = \frac{i\hbar}{2m} \left(\Psi \frac{\partial \Psi^*}{\partial x} - \Psi^* \frac{\partial \Psi}{\partial x} \right) \tag{1.1}$$

is defined such that

$$\frac{dP_{ab}(t)}{dt} = J(a, t) - J(b, t) \tag{1.2}$$

---

[1]It is also called the Landauer-Büttiker formalism. The first idea to relate the conductance of a small system to its scattering matrix comes from Landauer. Büttiker made a generalization to multi-terminal devices, which have more than two electrodes.

[2]Secs. 1.1 and 1.2 are adaptations from my lecture notes for the introductory bachelor course in quantum mechanics.

[3]The model is didactical and not designed to be the most efficient way to solve this scattering problem. The math is easy, however, and the general ideas can be applied to more complicated systems.

[4]Secs. 1.3-1.5 consist of elements of lecture notes I am writing for a master course.

describes the change in the probability $P_{ab}(t)$ of finding a particle in the region $a < x < b$ at time t, where

$$P_{ab}(t) = \int_a^b |\Psi(x,t)|^2\, dx, \tag{1.3}$$

provided the wave function $\Psi(x,t)$ describing the particle is normalized. Proving this is trivial:

$$
\begin{aligned}
\frac{dP_{ab}(t)}{dt} &= \int_a^b \left[\Psi^*(x,t)\frac{\partial \Psi(x,t)}{\partial t} + \frac{\partial \Psi^*(x,t)}{\partial t}\Psi(x,t) \right] dx \\
&= \frac{i\hbar}{2m} \int_a^b \left[\Psi^*(x,t)\frac{\partial^2 \Psi(x,t)}{\partial x^2} - \frac{\partial^2 \Psi^*(x,t)}{\partial x^2}\Psi(x,t) \right] dx \\
&= \frac{i\hbar}{2m} \left[\Psi^*(x,t)\frac{\partial \Psi(x,t)}{\partial x} - \frac{\partial \Psi^*(x,t)}{\partial x}\Psi(x,t) \right]_a^b.
\end{aligned}
$$

Going from the first to the second line one uses the Schrödinger equation $i\hbar\frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x)\Psi(x,t)$ and its complex conjugate (the potential terms cancel), and from the second to the third line one uses partial integration. Setting $b = a + dx$ with dx infinitesimal, allows one to write Eq. 1.2 as

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}, \tag{1.4}$$

with $\rho(x,t) = |\Psi(x,t)|^2$ the probability density. You might recognize Eq. 1.4 as a continuity equation, which describes the relation between a density and a current.

Probability currents may seem rather abstract, but they are easily related to something more familiar. Suppose the particle has a charge q, then the expected charge found in the region $a < x < b$ at time t is $Q_{ab}(t) = qP_{ab}(t)$.[5] Defining the **electrical current** as $I(x,t) = qJ(x,t)$, Eq. 1.2 can be rewritten as

$$\frac{dQ_{ab}(t)}{dt} = I(a,t) - I(b,t). \tag{1.5}$$

This makes sense; the rate of change of charge is given by the difference between the current flowing in from one side minus the current flowing out from the other side.

A nice thing is that, even if the *wave function cannot be normalized,* like the wave function of a free particle, the *probability current* according to Eq. 1.1 is still a *well-defined* quantity. Free particles often enter in scattering problems, where we are interested in quantities like reflection and transmission coefficients. Since the latter can be directly defined in terms of probability currents, we can get away with using non-normalizable wave functions.[6]

1.1.2   Stationary states

Suppose now that $\Psi(x,t)$ describes a **stationary state**, i.e.

$$\Psi(x,t) = \psi(x)e^{-\frac{i}{\hbar}Et}. \tag{1.6}$$

---

[5]This is an expectation value in the quantum mechanical sense. One starts the wave at time 0 and at time t one measures whether the particle is in the region $a < x < b$. By repeating this "experiment" over and over, one can calculate the probability $P_{ab}(t)$. $Q_{ab}(t)$ is then the average charge found in this region from these repeated "experiments". If you have problems imagining this then think of a particle emitter that sends out a pulse of many (independent) particles. The averaging is done automatically and the average charge is what you measure.

[6]For quantum purists: one *can* work with normalizable wave packets. The math then becomes ugly, and in the proper limit the physical results will be the same as when using plane waves.

Then one finds from Eq. 1.3

$$\frac{dP_{ab}(t)}{dt} = 0$$

and from Eqs. 1.2 and 1.1

$$J(x,t) = J = \quad \text{constant.} \tag{1.7}$$

The probability current is constant, i.e. independent of position and time.

For example, consider a free particle with the wave function

$$\psi(x) = A\, e^{ikx}. \tag{1.8}$$

From Eq. 1.3 we calculate

$$P_{ab} = |A|^2\, (b-a). \tag{1.9}$$

Since P_{ab} is the probability of finding the particle in the interval between $x = a$ and $x = b$, i.e. an interval of length $b - a$, we can interpret $|A|^2$ as the probability density per unit length. It is also called the particle density[7]

$$\rho = |A|^2. \tag{1.10}$$

The probability current is easily calculated from its definition, Eq. 1.1

$$J = \frac{\hbar k}{m}\,|A|^2 = \frac{\hbar k}{m}\rho. \tag{1.11}$$

According to de Broglie's relation $p = \hbar k$ is the momentum of the particle and

$$v = \frac{\hbar k}{m} = \frac{p}{m} \tag{1.12}$$

is then the velocity of the particle. The electrical current is given by

$$I = qJ = qv\rho, \tag{1.13}$$

which is the usual definition of an electrical current, namely charge×velocity×density. For the wave function of Eq. 1.8 both velocity and density are constant, so the wave function describes a uniform current. Suppose $q > 0$; then if $k > 0$ the current flows to the right, if $k < 0$, the current flows to the left. From now on we assume that $k > 0$.

Now let's go to the more complicated wave function

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \tag{1.14}$$

with A, B constants. The associated probability current is

$$J = \frac{\hbar k}{m}\,|A|^2 - \frac{\hbar k}{m}\,|B|^2, \tag{1.15}$$

which is interpreted as a right going current minus a left going current. In a scattering problem one would interpret the first term on the right handside of Eq. 1.14 as the incident wave and the second term as the reflected wave. Eq. 1.15 is then interpreted as the difference between incident and reflected currents

$$J = J_{in} - J_R. \tag{1.16}$$
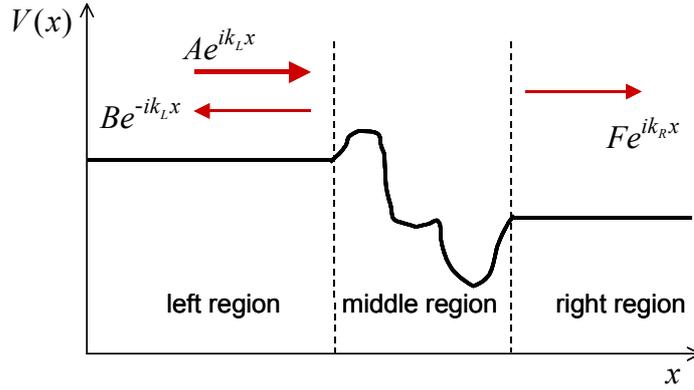
Figure 1.1: Cartoon representing a general one-dimenssional scattering problem. In the left region the potential is a constant $V(x) = V_L$, in the middle region the potential $V(x)$ can be anything, and in the right region the potential is a constant $V(x) = V_R$. The middle region is called the **scattering region**. The left and right regions are called the **left and right leads**. In the left lead we have an incoming wave Ae^{ik_Lx} and a reflected wave Be^{-ik_Lx} and in the right lead we have a transmitted wave Fe^{ik_Rx}.

The reflection coefficient R is defined as the ratio between reflected and incident currents

$$R = \frac{J_R}{J_{in}} = \frac{|B|^2}{|A|^2}. \tag{1.17}$$

Consider the scattering problem shown in Fig. 1.1. In the left region we assume that the potential is constant $V(x) = V_L$, in the middle region the potential can have any shape $V(x)$, and in the right region the potential is again constant $V(x) = V_R$. The solution in the left region is given by Eq. 1.14 with k replaced by k_L

$$k_L = \frac{\sqrt{2m(E - V_L)}}{\hbar}. \tag{1.18}$$

The solution in the right region is given by the transmitted wave

$$\psi(x) = Fe^{ik_Rx}\,;\; x \text{ in right region}, \tag{1.19}$$

with

$$k_R = \frac{\sqrt{2m(E - V_R)}}{\hbar}. \tag{1.20}$$

One can calculate the transmitted current as

$$J_T = \frac{\hbar k_R}{m}\,|F|^2\,. \tag{1.21}$$

The transmission coefficient T is defined as the ratio between transmitted and incident currents

$$T = \frac{J_T}{J_{in}} = \frac{v_R}{v_L}\frac{|F|^2}{|A|^2}, \tag{1.22}$$

---

[7]Using a beam of N particles in which each particle is independent of the others and is described by the wave function of Eq. 1.8, the particle density is $N|A|^2$, which is the number of particles to be found per unit length.

using Eq. 1.12.

From the fact that the current has to be independent of position *everywhere*, see Eq. 1.7, it follows from Eqs. 1.16 and 1.21 that

$$\begin{aligned} J_{in} - J_R &= J_T \Leftrightarrow \\ J_{in} &= J_R + J_T. \end{aligned} \tag{1.23}$$

This relation expresses the conservation of current, or: "current in = current out" (reflected plus transmitted). No matter how weird the potential in the middle region is, the current going into it has to be equal to the total current coming out of it. No particles magically appear or disappear in the middle region. From the definitions of Eqs. 1.17 and 1.22 it is shown that Eq. 1.23 is equivalent to

$$1 = R + T, \tag{1.24}$$

i.e. the reflection and transmission coefficients add up to 1. Since these coefficients denote the probabilities that a particle is reflected or transmitted, this simply states that particles are either reflected or transmitted.

1.2 Quantum conductance

1.2.1 Tunnel junction

The device shown in Fig. 1.2 is called a tunnel junction. The left and right regions consist of metals and the middle region consists of an insulator material, usually a metal-oxide.[8] Such devices can be made in a very controlled way with the middle region having a thickness of a few nm. One is interested in electrical currents, i.e. the transport of electrons through such junctions, or more generally in the current-voltage characteristics of such a device.[9] On this small, nanometer length scale electrons have to be considered as waves and quantum tunneling is important. **Nano-electronics** is the general name of the field where one designs and studies special devices that make use of this electron wave behavior.

We start with the simplest possible one-dimensional model of a tunnel junction. The atoms of a material attract electrons by their nuclear Coulomb potential. The electrons in low lying energy levels are localized around the atomic nuclei and form the atomic cores. The atomic valence electrons experience a much weaker effective atomic core potential, which is the sum of attractive nuclear and repulsive core electron terms. If the atoms are closely packed and the material is sufficiently simple, all these atomic potentials add up to a total potential that is relatively constant in space. As a qualitative level the potential for electrons in a material can be approximated by a constant, which is what we are going to do in the following.[10] The constant potential depends on the sort of atoms a material is composed of, so it is different for every material. The potential in the tunnel junction of Fig. 1.2 along the transport direction can then be represented by a square barrier, as shown in Fig. 1.3.

---

[8]Scanning tunneling microscopy (STM) uses a tunnel junction between the probe tip and a surface, where the middle region is simply vacuum.

[9]The device is called MIM, which stands for metal-insulator-metal. Using magnetic metals the device can be applied as a magnetic field sensor, or in MRAMs (magnetic random access memories).

[10]This approximation is often used for simple metals such as the alkali's, aluminium, silver and gold. The constant potential approximation is also called the **jellium approximation**. It does not hold for complicated metals such as the transition metals or for covalently bonded materials, such as silicon or carbon. To be fair, it doesn't even hold very well for simple metals if one is interested in quantitative results.
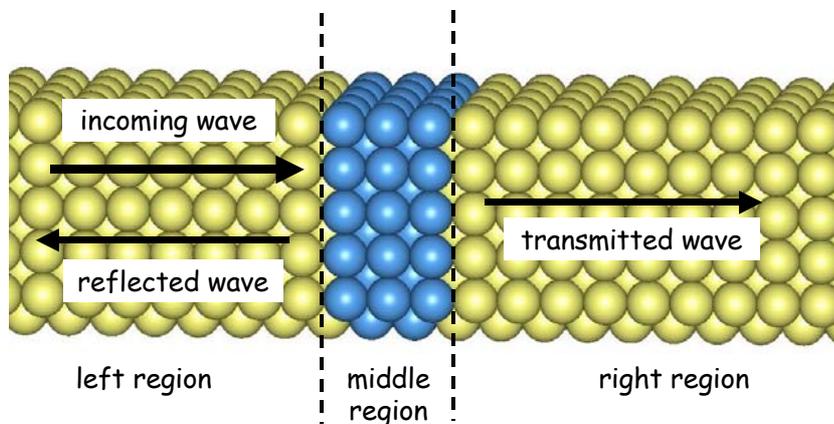
Figure 1.2: Schematic representation of a tunnel junction. The yellow balls represent atoms of a metal, the blue balls represent atoms of an insulator. The left and right regions stretch macroscopically far into the left and right, respectively. The electron waves in the metal are reflected or transmitted by the insulator in the middle region
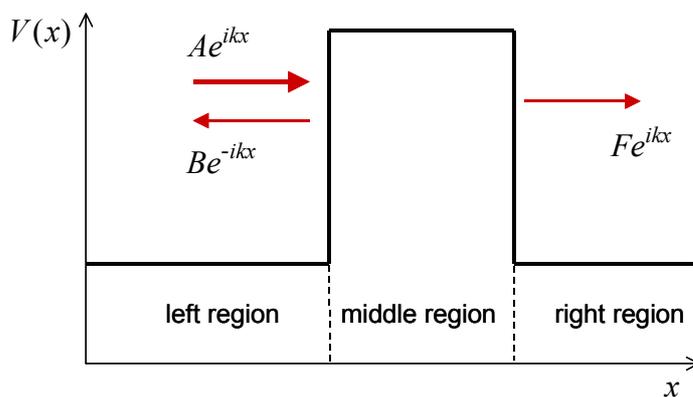
Figure 1.3: Simple approximation of the potential along the transport direction of a tunnel junction, see Fig. 1.2. In the metal (left and right regions) the potential is constant, $V(x) = V_1$. In the insulator the potential is also constant, $V(x) = V_0$, where $V_0 > V_1$. The incoming, reflected and transmitted waves are given by Ae^{ikx}, Be^{-ikx} and Fe^{ikx}.
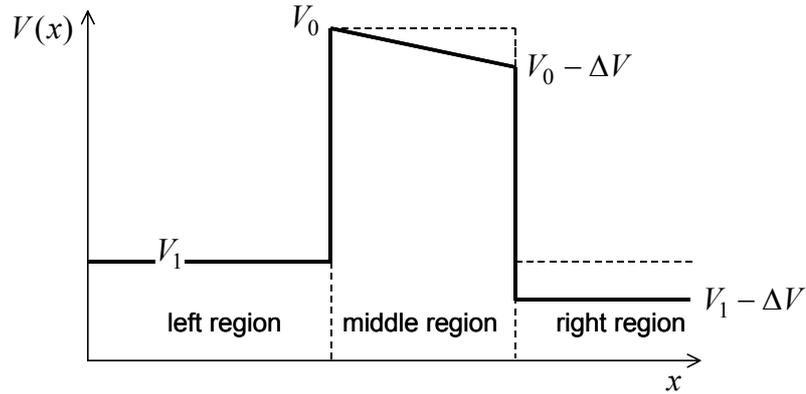
Figure 1.4: The potential when a bias voltage U is applied between the left and right leads. This changes the potential of the right region by $\Delta V = -eU$ to $V_1 - \Delta V$ with respect to the left region; see Fig. 1.3. The voltage drop is indicated schematically. If the bias voltage is small, i.e. $\Delta V \ll V_0 - V_1$, then we can still use the transmission coefficient T calculated for the unbiased square barrier (given by the dashed line).

1.2.2 The Landauer formula

According to Eq. 1.22 the transmitted electrical current is given by

$$I_T = I_{in}T, \tag{1.25}$$

using the definition $I = qJ$ (the charge q of an electron is $-e$). The incoming current I_{in} is given by Eq. 1.13 as

$$I_{in} = -ev\rho. \tag{1.26}$$

How large are the velocity v and density ρ of the incoming electrons? To answer that we must ask the more basic question: how is the incoming current in a device created? In an experimental setup this is done by applying a voltage difference U between the left and right regions. The left and right regions are metals, which can be connected to the two ends of a battery, for instance. This results in a potential drop $\Delta V = -eU$ between left and right regions, as shown in Fig. 1.4. We suppose that the temperature is zero and the metals are non-magnetic, so we have spin degeneracy. Then $v\rho$ is given by the simple expression

$$v\rho = \frac{\Delta V}{\pi\hbar} = -\frac{eU}{\pi\hbar}. \tag{1.27}$$

This expression is derived in the next section.

If the potential drop ΔV in Fig. 1.4 is small compared to the barrier height $V_0 - V_1$, we can use the unbiased square barrier potential from Fig. 1.3 to calculated the transmission coefficient T.[11] This is the so-called **linear response regime**. Eqs. 1.25-1.27 then give for the transmitted electrical current, also called the **tunneling current**

$$I_T = \frac{e^2}{\pi\hbar}U\,T, \tag{1.28}$$

---

[11] One can calculate the transmission coefficient T for the tilted barrier of Fig. 1.4, but this calculation is more complicated, and for small ΔV the result is almost the same as for the square barrier.

which is a remarkably simple expression! If we define the **conductance** $\mathcal{G}$ as current divided by voltage, we get

$$\mathcal{G} = \frac{I_T}{U} = \frac{e^2}{\pi\hbar}T. \tag{1.29}$$

Since T is just a dimensionless number between 0 and 1, $e^2/\pi\hbar$ has the dimension of conductance. It is the fundamental **quantum of conductance**; its value is $e^2/\pi\hbar \approx 7.75 \times 10^{-5}\ \Omega^{-1}$.[12]

Eq. 1.29 is called the **Landauer formula**; it plays a central role in nano-electronics.[13] As it stands here, it is valid for a one-dimensional, spin degenerate system at low voltage and at not too high a temperature, but it can be generalized.

1.2.3   Simple derivation of the Landauer formula

I will give a very simple derivation of Eq. 1.27. We have to do a little bit of solid state physics, but I use only simple introductory quantum mechanics language. Spin degeneracy means that each energy level can be filled with two electrons. The non spin degenerate case is relevant for magnetic materials. I let you work out that case yourself.

The Pauli exclusion principle and the Fermi energy

The left and right regions of a tunnel junction consist of metal wires, see Fig. 1.2. These wires are supposed to be very, very long compared to the size of the middle region. In a simple-minded model the potential of a metal wire looks like Fig. 1.5. The potential is approximately constant inside the wire and it has steps at the beginning and end of the wire to keep the electrons in. The energy levels of this square well potential are,[14]

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{n^2\pi^2\hbar^2}{2mL^2}. \tag{1.30}$$

The spacing between the energy levels, $E_n - E_{n-1}$, scales as $1/L^2$ with the length L of the wire. If L is large, the spacing becomes very small, so from a distance the energy level spectrum almost looks like a continuum, as illustrated by Fig. 1.5.

The wave functions are given by

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin k_n x = \frac{1}{i\sqrt{2L}} \left(e^{ik_n x} - e^{-ik_n x} \right). \tag{1.31}$$

These are not exactly what we need, because they correspond to standing waves, whereas we need traveling waves to describe currents, see Eq. 1.8. For the incoming current we only need the $\exp(ik_n x)$ part. Setting $A = 1/(i\sqrt{2L})$, the corresponding electron density according to Eq. 1.10 is

$$\rho = |A|^2 = \frac{1}{2L}. \tag{1.32}$$

---

[12]If you are more used to working with resistances, the resistance $\mathcal{R}$ is the inverse of the conductance, i.e. $\mathcal{R} = 1/\mathcal{G}$, so the quantum of resistance is $\pi\hbar/e^2 \approx 12.9\text{k}\Omega$.

[13]R. Landauer, *Philosophical Magazine* **21**, 863 (1970).

[14]This is actually the solution for an infinitely deep square well, whereas you might think that we need the solution for a finite square well. However if the well is very wide and not too shallow, the infinite well is an extremely good approximation.
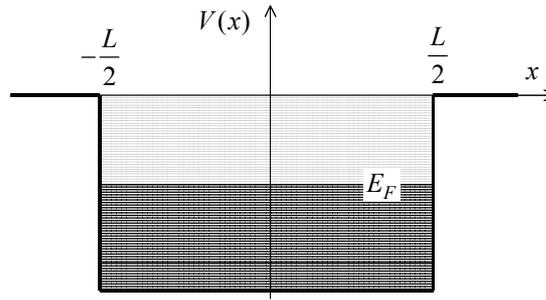
Figure 1.5: Schematic drawing of the potential and the energy levels of a long wire. The points $-L/2$ and $L/2$ mark the beginning and the end of the wire. The spacing between the energy levels is so small that the energy spectrum almost looks like a continuum. E_F marks the Fermi energy, i.e. the highest level that is occupied in the ground state by an electron.

The wire is full of electrons since each of the atoms in the wire brings at least one electron with it. Filling the energy levels according to the Pauli principle, and having N electrons in total, the highest occupied level is $E_{\frac{N}{2}}$. The highest occupied level in the ground state is called the **Fermi energy** or E_F.[15]

Incoming and tunnel currents

Go back to the tunnel junction and fill its potential profile, Fig. 1.3, with electrons from the left and right wires. This is shown in Fig. 1.6. The Fermi energy E_F in the left and right regions is the same. The exclusion principle then tells us that there can be no flow of current. Any electron on the left side that would try to go to the right side finds an energy level that is already occupied by an electron, which excludes any other electron from going there, and vice versa. This confirms what we know from everyday life; in an unbiased system no current flows.[16]

Now apply a bias voltage between left and right regions as in Fig. 1.4. The result is shown in Fig. 1.7. The bias voltage lowers the right region in energy. Suddenly the electrons on the left side that occupy energy levels E_n with $E_F - \Delta V < E_n < E_F$ find empty levels with that energy on the right side. They can tunnel through the barrier to occupy these levels. Provided the potential drop ΔV is small, we can approximate the transmission coefficients of all these electrons by T at an energy $E = E_F$.[17] The incoming current of Eq. 1.26 has contributions from _all_ electrons with energies between $E_F - \Delta V$ and E_F.

$$I_{in} = -2e\rho \sum_{E_F - \Delta V < E_n < E_F} v_n. \tag{1.33}$$

The factor of 2 is there because there are two electrons in each level.

---

[15]In an ordinary metal the Fermi energy is of order 10 eV with respect to the lowest energy level of the valence electrons.

[16]One can also reverse the argument. If the Fermi energies on the left and right side would be different, then a current would flow. However this current would be short-lived. By sending electrons from left to right one occupies a level on the right side, and de-occupies a level on the left side. This would go on until the highest occupied levels on the left and right are the same; in other words until an equilibrium is reached.

[17]Again, this is valid in the linear response regime.
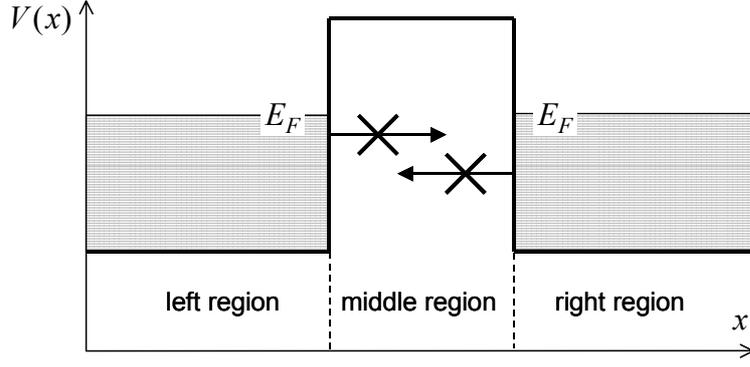
Figure 1.6: Tunnel junction where left and right regions are filled with electrons. The Fermi energies E_F on the left and right side are identical. The exclusion principle forbids electrons to trespass from left to right or vice versa.

This sum in Eq. 1.33 is rather awkward, but by a trick we can turning it into an integral

$$\sum v_n = \frac{L}{\pi} \sum v_n \frac{\pi}{L} = \frac{L}{\pi} \sum v_n \Delta k$$
$$\approx \frac{L}{\pi} \int v dk, \tag{1.34}$$

where

$$\Delta k = k_n - k_{n-1} = \frac{\pi}{L}, \tag{1.35}$$

see Eq. 1.30. Turning the sum into an integral is allowed because L is very large, so Δk is tiny. The lower and upper bound of the integral in Eq. 1.34 should correspond to the energies $E_F - \Delta V$ and E_F, whereas the integral is over dk, which is again awkward. We can however turn it into an integral over dE, using the following trick

$$v = \frac{\hbar k}{m} = \frac{1}{\hbar} \frac{d \left(\frac{\hbar^2 k^2}{2m} \right)}{dk} = \frac{1}{\hbar} \frac{dE}{dk}. \tag{1.36}$$

In Eq. 1.34 it gives

$$\int_{E_F - \Delta V}^{E_F} v dk = \int_{E_F - \Delta V}^{E_F} \frac{1}{\hbar} \frac{dE}{dk} dk = \int_{E_F - \Delta V}^{E_F} \frac{1}{\hbar} dE$$
$$= \frac{1}{\hbar} \left(E_F - (E_F - \Delta V) \right) = \frac{1}{\hbar} \Delta V. \tag{1.37}$$

Collecting Eqs. 1.37, 1.34 and Eq. 1.32 in Eq. 1.33, we find for the incoming current

$$I_{in} = -e \frac{\Delta V}{\pi \hbar}. \tag{1.38}$$

This is the required expression for the incoming current, see Eqs. 1.26 and 1.27. The tunnel current is then given by

$$I_T = I_{in} T = -\frac{e \Delta V}{\pi \hbar} T, \tag{1.39}$$

where the transmission coefficient T needs to be calculated for the energy $E = E_F$. Note that with $\Delta V = -eU$, where U is the potential difference (in Volts), this corresponds to Eq. 1.28. The Landauer formula, Eq. 1.29, is then derived straightforwardly.
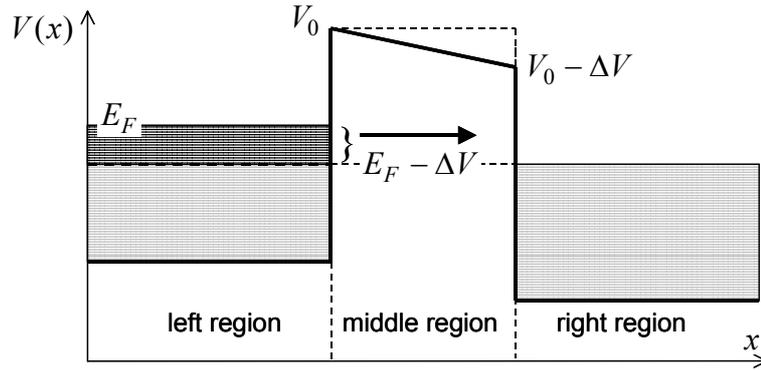
Figure 1.7: Tunnel junction with an applied bias voltage. All the levels occupied by electrons in the left region with an energy $E_F - \Delta V < E_n < E_F$ correspond to empty energy levels in the right region. The electrons in these levels can tunnel from though the barrier from left to right.

1.3 Transfer and scattering matrices

The Landauer formula expresses the conductance in terms of a transmission coefficient. In other words, the problem of finding the conductance becomes a problem of solving the scattering problem. In this section I review some basic elements of scattering theory in one dimension. You probably know what a scattering matrix is, but read through the section anyway, if only to refresh your memory.

1.3.1 Transmission of a rectangular barrier

Again I explain the concepts using elementary quantum mechanics only and the rectangular barrier as an example. Consider Fig. 1.8 and let the middle region run from $x = -a$ to $x = a$. We are trying to find a solution to the time-independent Schrödinger equation

$$E\psi(x) + \frac{\hbar^2}{2m} \frac{d^2\psi(x)}{dx^2} - V(x)\psi(x) = 0. \tag{1.40}$$

Since the potential is piecewise constant, we can write the solution as

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx}; & x < -a \\ Ce^{\eta x} + De^{-\eta x}; & -a < x < a \\ Fe^{ikx} + Ge^{-ikx}; & x > a \end{cases} \tag{1.41}$$

with

$$k = \sqrt{\frac{2mE}{\hbar}}, \ \eta = \sqrt{\frac{2m(V_0 - E)}{\hbar}}. \tag{1.42}$$

For $0 < E < V_0$, both k and η are real numbers.[18] Provided we choose the constants A-G such, that the function $\psi(x)$ is continuous and differentiable everywhere, it will be a solution to the Schrödinger equation for all x, including the boundaries $x = \pm a$. Continuity of $\psi(x)$ at $x = -a$ gives the relation

$$Ae^{-ika} + Be^{ika} = Ce^{-\eta a} + De^{\eta a},$$

---

[18]Most of the equations actually also hold for $0 < V_0 < E$ (scattering over the barrier) or for $V_0 < 0 < E$ (scattering from a potential well). In those cases η will be a purely imaginary number.
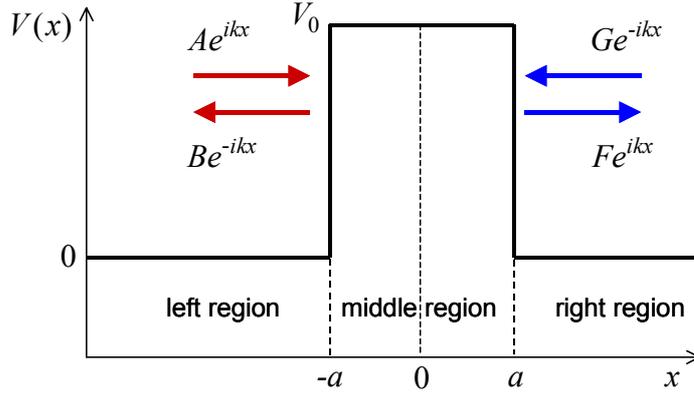
Figure 1.8: Scattering from a rextangular barrier. The transfer matrix $\mathbf{M}$ relates the coefficients of the waves at the right to the waves at the left of the barrier.

whereas continuity of $\frac{d}{dx}\psi(x)$ at $x = -a$ gives

$$ikAe^{-ika} - ikBe^{ika} = \eta Ce^{-\eta a} - \eta De^{\eta a}.$$

These two equations can be combined into the matrix equation

$$\begin{pmatrix} A \\ B \end{pmatrix} = \mathbf{M}^{(1)} \begin{pmatrix} C \\ D \end{pmatrix} \tag{1.43}$$

$$\mathbf{M}^{(1)} = \begin{pmatrix} \left(\frac{ik+\eta}{2ik}\right)e^{(ik-\eta)a} & \left(\frac{ik-\eta}{2ik}\right)e^{(ik+\eta)a} \\ \left(\frac{ik-\eta}{2ik}\right)e^{-(ik+\eta)a} & \left(\frac{ik+\eta}{2ik}\right)e^{-(ik-\eta)a} \end{pmatrix}.$$

In a similar way, continuity of $\psi(x)$ and $\frac{d}{dx}\psi(x)$ at $x = a$ gives the matrix equation

$$\begin{pmatrix} C \\ D \end{pmatrix} = \mathbf{M}^{(2)} \begin{pmatrix} F \\ G \end{pmatrix} \tag{1.44}$$

where the matrix $\mathbf{M}^{(2)}$ can be obtained from $\mathbf{M}^{(1)}$ by replacing $-a$ by a, ik by η and η by ik. Combining Eqs. 1.43 and 1.44 gives an equation of the form

$$\begin{pmatrix} A \\ B \end{pmatrix} = \mathbf{M} \begin{pmatrix} F \\ G \end{pmatrix} \tag{1.45}$$

$$\mathbf{M} = \mathbf{M}^{(1)}\mathbf{M}^{(2)},$$

where the 2×2 matrix $\mathbf{M}$ is the product of the two matrices of Eqs. 1.43 and 1.44.

In a scattering problem we set our boundary conditions by hand, or, if you wish, by choosing specific experimental conditions. We define an incoming wave Ae^{ikx}, with A a fixed value that gives the density of particles corresponding to our experimental source, see Eq. 1.10. At the same time we set $G = 0$, which means that we have no incoming wave Ge^{-ikx} from the right (experimentally this is easy, even for a theoretician). Comparing Eqs. 1.22 and 1.45 we find for the transmission coefficient

$$T = \frac{v_R}{v_L}\left|\frac{1}{M_{11}}\right|^2. \tag{1.46}$$
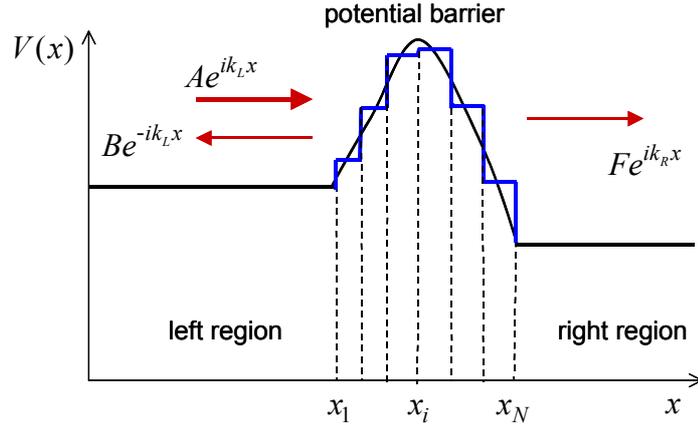
Figure 1.9: Approximating a barrier of any shape by a series of steps. The transfer matrix $\mathbf{M}$ is given by $\mathbf{M} = \mathbf{M}_1 \mathbf{M}_1 \cdots \mathbf{M}_N$.

Since the potential is symmetric, we actually have $v_R = v_L$ here, but we let the more general form of Eq. 1.46 stand for later. It is a simple exercise to work out the matrix element M_{11} and show that one obtains the usual textbook expression for the transmission coefficient

$$T = \left[1 + \left(\frac{k^2 + \eta^2}{2k\eta}\right)\sinh^2(2\eta a)\right]^{-1}. \tag{1.47}$$

Comparing Eqs. 1.17 and 1.45 gives the reflection coefficient as

$$R = \frac{|M_{21}|^2}{|M_{11}|^2}. \tag{1.48}$$

1.3.2   Transmission of a general barrier

The algorithm formulated in the previous section can be extended in order to describe the transmission through a barrier of a more general shape. The matrix $\mathbf{M}$ as in Eq. 1.45 is called the **transfer matrix**. It relates the modes in the right region to the modes in the left region. Once we know the transfer matrix, we can calculate the reflection and transmission coefficients, see Eqs. 1.48 and 1.46. For a single step in the potential, the transfer matrices are given by Eqs. 1.43 and 1.44 for a step up and a step down, respectively. Since the square barrier can be viewed as a step up (at $x = -a$), followed by step down (at $x = a$), its transfer matrix is simply a product of the transfer matrices of the individual steps, as is expressed by Eq. 1.45. The idea can be applied to a potential of any shape, approximating the potential by a series of steps, as is illustrated in Fig. 1.9.

The transfer matrix $\mathbf{M}$ is given by the product of the transfer matrices of each step

$$\mathbf{M} = \mathbf{M}_1 \mathbf{M}_1 \cdots \mathbf{M}_N, \tag{1.49}$$

where each $\mathbf{M}_i$ can be calculated by defining

$$\eta_i = \frac{\sqrt{2m\left[V(x_i) - E\right]}}{\hbar}, \tag{1.50}$$

and using a properly modified Eq. 1.43 or Eq. 1.44. I leave the details up to you. In essence this is the **transfer matrix algorithm**. It is used quite frequently in numerical calculations solving

scattering problems in various branches of physics; quantum mechanics, optics, acoustics, etc.. Obviously it is best suited for layered materials, i.e. systems that consist of layers of different materials stacked on top of one another. If the η_i get large, which is for high barriers and/or low energies, one might get into numerical problems because of the exponentials in Eqs. 1.43 and 1.44. There are some special tricks to handle these, but I won't go into details here. For systems described in terms of real atoms, the techniques discussed in the following sections are better suited.

1.3.3 The scattering or S matrix

The transfer matrix $\mathbf{M}$ is closely related to an elementary concept in scattering theory, called the scattering matrix. Whereas the transfer matrix relates the modes on the right to the modes on the left, the **scattering matrix** relates outgoing modes to incoming modes. For the example shown in Fig. 1.8, the modes that are coming into the barrier are Ae^{ikx} (from the left) and Ge^{-ikx} (from the right) and the modes that are going out from the barrier are Be^{-ikx} (to the left) and Fe^{ikx} (to the right). The scattering matrix is defined by the relation

$$\begin{pmatrix} B \\ F \end{pmatrix} = \mathbf{S}' \begin{pmatrix} A \\ G \end{pmatrix}. \tag{1.51}$$

Comparison to Eq. 1.45 gives for its matrix elements

$$\begin{aligned} S'_{11} &= \frac{M_{21}}{M_{11}}; \quad S'_{12} = M_{22} - M_{21}\frac{M_{12}}{M_{11}}; \\ S'_{21} &= \frac{1}{M_{11}}; \quad S'_{22} = -\frac{M_{12}}{M_{11}}. \end{aligned} \tag{1.52}$$

I have put a "prime" on $\mathbf{S}'$ because the scattering matrix $\mathbf{S}$ as it ordinarily used is defined as

$$\begin{aligned} S_{11} &= S'_{11}; \quad S_{12} = \sqrt{\frac{v_L}{v_R}}S'_{12}; \\ S_{21} &= \sqrt{\frac{v_R}{v_L}}S'_{21}; \quad S_{22} = S'_{22}. \end{aligned} \tag{1.53}$$

The reason for introducing the $\sqrt{v}$ factors is because one wants the scattering matrix $\mathbf{S}$ to reflect a basic conservation law, namely the conservation of current. In a stationary problem the current going out must be the same as the current coming in, since otherwise one would get an accumulation or depletion of particles, as discussed Sec. 1.1.

$$\begin{aligned} J_{out} &= J_{in} \Rightarrow \\ v_L \left|B\right|^2 + v_R \left|F\right|^2 &= v_L \left|A\right|^2 + v_R \left|G\right|^2 \Rightarrow \\ \left\| \begin{pmatrix} \sqrt{v_L}B \\ \sqrt{v_R}F \end{pmatrix} \right\|^2 &= \left\| \begin{pmatrix} \sqrt{v_L}A \\ \sqrt{v_R}G \end{pmatrix} \right\|^2. \end{aligned} \tag{1.54}$$

From Eqs. 1.51 and 1.53 it is easily shown that this holds only if

$$\mathbf{S}^\dagger\mathbf{S} = \mathbf{S}\mathbf{S}^\dagger = \mathbf{1}. \tag{1.55}$$

In other words, **conservation of current** means that the **scattering matrix is unitary**.

It is custom to write the scattering matrix as

$$\mathbf{S} = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix}. \tag{1.56}$$
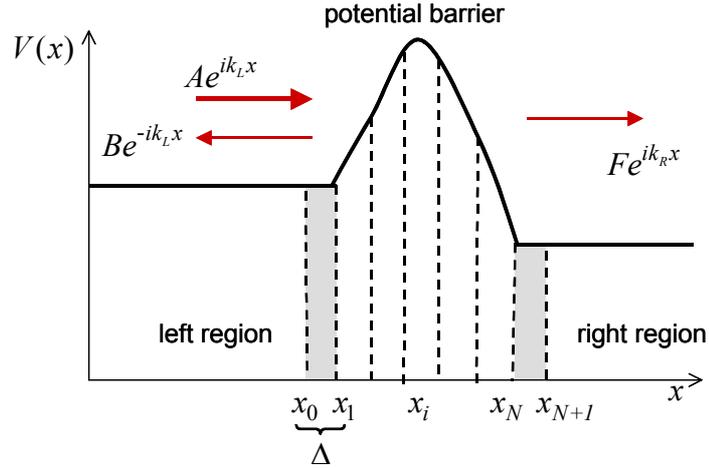
Figure 1.10: Discretization of a potential barrier. An equidistant grid is applied with $\Delta = x_{i+1} - x_i$. The potential is different from a constant for $x_1 \leq x \leq x_N$. The coupling to the left and right regions (the leads) can be included by extending the grid by two points, which defines the boundary zones, indicated in gray.

From Eqs. 1.46, 1.48, 1.52 and 1.53 it follows that the transmission and reflection coefficients are given by

$$T = |t|^2 ; \quad R = |r|^2 , \tag{1.57}$$

hence the name transmission and reflection amplitudes for t and r, respectively. Note that the relation $1 = R + T$, see Eq. 1.24, then simply reflects the unitarity of the scattering matrix.

1.4 Mode matching

Although the transfer matrix algorithm is quite general, it becomes quite tedious if the potential in the leads is not constant, which is when we want to make an atomistic model of a tunnel junction, see Fig. 1.2. In this section I will discuss a technique that is easier to generalize, called **mode matching**. By means of an introduction, I will discuss this technique in its simplest form and take Fig. 1.10 as starting point. The potential is discretized as before, but the grid is extended by one point into the left and right regions.

The basic idea is to discretize the whole Schrödinger equation, Eq. 1.40, including the kinetic energy. Approximating the second derivative by a simple, first order finite difference one obtains

$$E\psi_i + \frac{\hbar^2}{2m} \left\{ \frac{(\psi_{i+1} - \psi_i) - (\psi_i - \psi_{i-1})}{\Delta^2} \right\} - V_i\psi_i = 0, \tag{1.58}$$

where ψ_i and V_i are shorthand notations for $\psi(x_i)$ and $V(x_i)$, and $\Delta = x_{i+1} - x_i$. I only consider equidistant grids here. As is usual in a scattering problem, i runs from $-\infty$ to ∞, so we have an infinite number of equations. This is not only awkward, but also unnecessary. The potential is localized in space, i.e. $V(x)$ differs from a constant only for $x_1 \leq x \leq x_N$. For the left and right regions, $x < x_1$ and $x > x_N$, we already know the solutions to the Schrödinger equation. They are simple plane waves, as indicated in Fig. 1.10, with

$$k_L = \frac{\sqrt{2m(E - V_L)}}{\hbar} ; \quad k_R = \frac{\sqrt{2m(E - V_R)}}{\hbar}. \tag{1.59}$$

We need a way of matching these "modes" to the wave function in the region of the potential barrier (the scattering region).

Let's start with x_0, and put the origin there, so $x_0 = 0$. The finite difference Schrödinger equation for $i = 0$ is

$$E\psi_0 + \frac{\hbar^2}{2m\Delta^2}\left\{\psi_1 - 2\psi_0 + \psi_{-1}\right\} - V_0\psi_0 = 0. \tag{1.60}$$

We know that for $x < 0$ the wave function has the form $\psi(x) = Ae^{ik_L x} + Be^{-ik_L x}$, so

$$\psi_{-1} = Ae^{-ik_L\Delta} + Be^{ik_L\Delta} = Ae^{-ik_L\Delta} + (\psi_0 - A)e^{ik_L\Delta}, \tag{1.61}$$

where the last result follows from the fact that the wave function has to be continuous at $x = 0$. Remember that in a scattering problem we assume that we know the incoming wave, so A is fixed. Eq. 1.60 can then be rewritten as

$$E\psi_0 + \frac{\hbar^2}{2m\Delta^2}\left\{\psi_1 - 2\psi_0 + e^{ik_L\Delta}\psi_0\right\} - V_0\psi_0 = \frac{\hbar^2}{2m\Delta^2}A\left\{e^{ik_L\Delta} - e^{-ik_L\Delta}\right\}. \tag{1.62}$$

The term on the left handside now only contains ψ_i with $i \geq 0$, and the terms at the right handside can be considered as the source of the incoming wave. This takes care of the left boundary.

Now focus upon the right boundary; for $i = N + 1$ we have

$$E\psi_{N+1} + \frac{\hbar^2}{2m\Delta^2}\left\{\psi_{N+2} - 2\psi_{N+1} + \psi_N\right\} - V_{N+1}\psi_{N+1} = 0. \tag{1.63}$$

We now use

$$\psi_{N+2} = Fe^{ik_R(N+2)\Delta} = \psi_{N+1}e^{ik_R\Delta} \tag{1.64}$$

to get

$$E\psi_{N+1} + \frac{\hbar^2}{2m\Delta^2}\left\{\psi_{N+1}e^{ik_R\Delta} - 2\psi_{N+1} + \psi_N\right\} - V_{N+1}\psi_{N+1} = 0. \tag{1.65}$$

Note that we have again used our knowledge of the scattering boundary conditions. We have assumed only a transmitted wave in Eq. 1.64 (no incoming wave from this side) and wave function continuity at the boundary. The term on the left handside of Eq. 1.65 only contains ψ_i with $i \leq N + 1$. This takes care of the right boundary.

For $i = 1, \ldots, N$ we have no problems and we can use Eq. 1.58. We can collect these equations together with Eqs. 1.62 and 1.65 and summarize the problem as

$$\left(E\mathbf{I} - \mathbf{H}'\right)\boldsymbol{\psi} = \mathbf{q}. \tag{1.66}$$

Here $\boldsymbol{\psi}$ is a vector containing the coefficients $\psi_i; i = 0, \ldots, N + 1$, $\mathbf{q}$ is the "source" vector of length $N + 1$, where all the coefficients are zero, except the first one

$$q_0 = \frac{\hbar^2}{2m\Delta^2}A\left\{e^{ik\Delta} - e^{-ik\Delta}\right\}, \tag{1.67}$$

and $\mathbf{H}'$ is a $(N + 2) \times (N + 2)$ Hamiltonian matrix. All off-diagonal matrix elements are zero, except the sub- and super-diagonal ones

$$H'_{i,i+1} = H'_{i,i-1} = -\frac{\hbar^2}{2m\Delta^2}. \tag{1.68}$$

All diagonal matrix elements are identical to that of the original finite difference Hamiltonian

$$H'_{i,i} = -\frac{\hbar^2}{m\Delta^2} + V_i, \tag{1.69}$$

except the first and the last one, which are modified to

$$
\begin{aligned}
H'_{0,0} &= -\frac{\hbar^2}{m\Delta^2} + V_0 + \Sigma_L(E), \\
H'_{N+1,N+1} &= -\frac{\hbar^2}{m\Delta^2} + V_{N+1} + \Sigma_R(E).
\end{aligned}
\tag{1.70}
$$

with

$$
\begin{aligned}
\Sigma_L(E) &= -\frac{\hbar^2}{2m\Delta^2} e^{ik_L\Delta}, \\
\Sigma_R(E) &= -\frac{\hbar^2}{2m\Delta^2} e^{ik_R\Delta}.
\end{aligned}
\tag{1.71}
$$

The quantities $\Sigma_{L/R}(E)$ are called the **self-energies** of the left and right leads.[19] They take care of the proper coupling of the potential barrier to the outer regions, and contain all the information we require about the outer regions. The self-energy depends upon the energy of the incoming and scattered waves, see Eq. 1.59. Note that, while Eq. 1.58 represents an infinite dimensional problem, by introducing the self-energy (and the source), we have reduced it to a **finite, $N+2$ dimensional problem**, Eq. 1.66. That can be solved using standard algorithms for solving linear equations.[20]

Once you have solved Eq. 1.66, the only thing remaining is to extract the transmission and reflection amplitudes. The transmission amplitude is simply given by the wave function at the right side of the barrier, normalized to the incoming wave, and normalized with the velocities (to attain a unitary scattering matrix, see Eq. 1.53)

$$t = \sqrt{\frac{v_R}{v_L}} \frac{\psi_{N+1}}{A}. \tag{1.72}$$

The reflection amplitude is similarly determined from the wave function on the left side minus the incoming wave, normalized to the incoming wave

$$r = \frac{\psi_0 - A}{A}. \tag{1.73}$$

Some care should be taken in determining the velocities. Since we have discretized the Schrödinger equation, it is consistent to discretize the expression for the current, Eq. 1.1, in a similar way

$$J = \frac{i\hbar}{2m} \left(\psi_i \frac{\psi^*_{i+1} - \psi^*_i}{\Delta} - \psi^*_i \frac{\psi_{i+1} - \psi_i}{\Delta} \right). \tag{1.74}$$

This should actually give a position independent result, since we are considering a stationary problem. For a simple plane wave Ae^{ikx} this expression gives

$$J = \frac{i\hbar |A|^2}{2m\Delta} \left(e^{-ik\Delta} - e^{ik\Delta} \right),$$

---

[19]a somewhat unfortunate term descending from Green function folklore.
[20]In this case it is particularly simple, since the matrix is tridiagonal.

and comparison to Eqs. 1.11 and 1.12 gives

$$v = \frac{i\hbar}{2m\Delta}\left(e^{-ik\Delta} - e^{ik\Delta}\right). \tag{1.75}$$

Note that the continuum limit $\lim_{\Delta\to 0}$ of this expression gives Eq. 1.12.

The source term, Eq. 1.67, can then be simplified to

$$q_0 = \frac{i\hbar A}{\Delta}v_L. \tag{1.76}$$

In addition, from Eqs. 1.75 and 1.71 one can relate the velocity to the self-energy

$$v_{L/R} = -\frac{2\Delta}{\hbar}\operatorname{Im}\Sigma_{L/R}(E). \tag{1.77}$$

1.4.1   Green function expressions

Using Green functions the mode matching results can be put into a very compact, albeit somewhat obscure, form. Define a matrix

$$\mathbf{G}(E) = \left(E\mathbf{I} - \mathbf{H}'\right)^{-1}. \tag{1.78}$$

It is called a **Green-function matrix**. Note that it has a dimension $N+2$, as has the modified Hamiltonian matrix $\mathbf{H}'$. One can also define the infinite dimensional retarded Green-function matrix related to the original infinite dimensional Hamiltonian

$$\mathbf{G}^r(E) = [(E + i\eta)\mathbf{I} - \mathbf{H}]^{-1}, \tag{1.79}$$

where η is (real, positive) infinitesimal. The advanced Green-function matrix is defined as[21]

$$\mathbf{G}^a(E) = [\mathbf{G}^r(E)]^\dagger. \tag{1.80}$$

For z a complex number in the lower half plane, the matrix elements of $\mathbf{G}$ and $\mathbf{G}^r$ in the scattering region are identical.

$$G_{i,j}(z) = G_{i,j}^r(z) \; ; \; i,j = 0,\dots,N+1 \tag{1.81}$$

A proof of this can be found in the literature. Note that the modified Hamiltonian matrix $\mathbf{H}'$ is non-Hermitian, essentially because the self-energy Σ is not real, see Eqs. 1.70 and 1.71. One can show that the eigenvalues of $\mathbf{H}'$ are not real and lie in the upper half complex plane. Thus, $\mathbf{G}(E)$ is a well-defined quantity for real energies E (unlike $\mathbf{G}^r$). It has the retarded boundary condition build into it and one does not need the $+i\eta$ trick.[22]

The definition of $\mathbf{G}$ allows us to write,

$$\psi_{N+1} = G_{N+1,0}(E)q_0, \tag{1.82}$$

---

[21]Since E is an eigenvalue of $\mathbf{H}$, $E\mathbf{I} - \mathbf{H}$ is singular and its inverse does not exist for real energies E. Adding/subtracting an imaginary $i\eta$ avoids the singularities. In textbooks on scattering theory it is shown that this has a physical meaning. $\mathbf{G}^r$ can be used to construct the retarded wave function, which consists of a wave coming in from the source to the target, and waves scattered out from the target. This is the physical solution. $\mathbf{G}^a$ gives the reverse, i.e. waves scattered into the target, plus waves going into the source. This is unphysical, but can sometimes be useful in formal mathematical manipulations. After having constructed the wave function, the limit $\lim_{\eta\to 0}$ can be taken, so η only serves as an intermediate to reflect the boundary conditions.

[22]Mathematicians would call $\mathbf{G}(E)$ an analytical continuation of $\mathbf{G}^r(E)$. All the poles of $\mathbf{G}(E)$ are in the upper half plane, which makes it analytical on the real axis, and a retarded form.

see Eq. 1.66. Eqs. 1.72, 1.76 and 1.82 then lead to a compact expression for the transmission amplitude

$$t = \frac{i\hbar}{\Delta} \sqrt{v_R v_L} \; G_{N+1,0}(E).$$ (1.83)

An expression of this type is called a **Fisher-Lee expression**. It relates matrix elements of the scattering matrix to matrix elements of the Green function.

For Green function junkies, we can even make it fancier and write, using Eq. 1.77

$$t = 2i\sqrt{-\operatorname{Im}\Sigma_R} \; G_{N+1,0}(E) \; \sqrt{-\operatorname{Im}\Sigma_L}.$$ (1.84)

This allows the transmission coefficient to be written as

$$T = t^*t = 4(\operatorname{Im}\Sigma_R)G^r_{N+1,0}(\operatorname{Im}\Sigma_L)G^a_{0,N+1}$$ (1.85)

with all quantities evaluated at the fixed energy E. This expression is known as the **Caroli expression**.

Introducing a Green function to tackle this simple one-dimensional problem is like using a sledgehammer to crack a nut. Green functions expressions can however be modified to include multiple dimensions, a large bias voltage, interactions with vibrations and/or between electrons, etc., where they give compact (though not necessarily practical) expressions. In case of a large bias voltage, the Caroli expression is also known as the **NEGF expression**, where NEGF stands for non-equilibrium Green function.[23]

1.4.2 The embedding potential

If you prefer differential equations over linear algebra, we can take the continuum limit of Eq. 1.66, i.e. $\lim_{\Delta\to0}$. It is tricky, but straightforward to obtain

$$E\psi(x) + \frac{\hbar^2}{2m}\left(\frac{d^2\psi(x)}{dx^2} + \delta(x)\frac{d\psi(x)}{dx} - \delta(x-a)\frac{d\psi(x)}{dx}\right)$$

$$-V(x)\psi(x) - \{\delta(x)\Sigma_L(E) + \delta(x-a)\Sigma_R(E)\}\,\psi(x) = i\hbar v_L A\delta(x),$$ (1.86)

where I have put the left boundary at $x = 0$ and the right boundary at $x = a$. The Σ's have the form

$$\Sigma_{L/R}(E) = -\frac{i\hbar v_{L/R}}{2}.$$ (1.87)

The term on the right handside of Eq. 1.86 describes the source term, which only exists at the left boundary. On the left handside there are δ-function terms that only exists at the boundaries of the barrier with the outside regions. They take care of the coupling to the outside region. The $\frac{d}{dx}$ terms in the kinetic energy ensure a continuous derivative across the boundaries. The potential term between { } is called the **embedding potential**. The formalism is known in the literature as the embedding formalism.[24] The differential equation needs to be solved for $\psi(x)$, but only over the finite domain $0 \le x \le a$. The coupling to the outside regions is taken care of by the boundary terms. Eqs. 1.72 and 1.73 give the transmission and reflection amplitudes with $\psi(a)$ replacing ψ_{N+1} and $\psi(0)$ replacing ψ_0.

---

[23]This is in contrast to the linear response regime, where the Green functions are ordinary, i.e. equilibrium Green functions.

[24]Be careful however, as in different fields the phrase "embedding" is attached to different things.

Alternatively, one may use Green function expressions also in this case. Define the Green function $G(x, x', E)$ as the solution of the equation

$$\left(E - \widehat{H}'\right) G(x, x', E) = \delta(x - x'),$$

(1.88)

where $\widehat{H}'$ is the operator of Eq. 1.86 including all potential and boundary terms. One can prove that $G(x, x', E)$ corresponds to the usual retarded Green function G^r for $0 \leq x, x' \leq a$.[25] The continuum equivalents of Eqs. 1.82-1.85 are

$$
\begin{aligned}
\psi(a) &= G(a, 0, E) i \hbar v_L A, \\
t &= i \hbar \sqrt{v_R v_L}\, G(a, 0, E) = 2i \sqrt{-\operatorname{Im}\Sigma_R}\, G(a, 0, E)\, \sqrt{-\operatorname{Im}\Sigma_L}, \\
T &= 4 \operatorname{Im}\Sigma_R G^r(a, 0, E) \operatorname{Im}\Sigma_L G^a(0, a, E).
\end{aligned}
$$

(1.89)

I prefer linear algebra over differential equations and I don't have experience with the embedding formalism, so it won't be discussed further. It has been proven however that the formalism can be extended into a practical method for calculating transport through small systems, including all atomic details.

1.5   Tight-binding

In previous sections we have analyzed one-dimensional scattering problems starting from potentials and wave functions as functions of a continuous coordinate. We have discretized this representation to show how the scattering problem can be solved in a practical way. This approach is natural if the potential variation is confined to the scattering region and outside this region the potential is constant. The incoming, reflected and transmitted waves are then simply plane waves. With the tunnel junction in mind, see Fig. 1.2, we would like to consider the situation in which the whole space is filled with atoms of one kind or another. In one dimension this gives an atomic wire, with atomic potentials everywhere along the wire, as is illustrated in Fig. 1.11.

The usual approach is to construct a representation on a basis of atomic orbitals, i.e. expand the wave functions in fixed atomic orbitals $\chi_i(x)$

$$\psi(x) = \sum_i c_i \chi_i(x - X_i),$$

(1.90)

where X_i denote the positions of the atoms. In chemistry this is known as the LCAO representation (linear combination of atomic orbitals), in physics it is usually called the **tight-binding representation**. The wave function is represented by the column vector of the coefficients

$$
\boldsymbol{\psi} = \begin{pmatrix} \vdots \\ c_{i-1} \\ c_i \\ c_{i+1} \\ \vdots \end{pmatrix}.
$$

(1.91)

Since we want to solve a scattering problem, the atomic orbitals should cover all of (one-dimensional) space, so the vector has an infinite dimension, i.e. $i = -\infty, \ldots, \infty$. Operators $\widehat{A}$

---

[25] $G^r(x, x', E)$ is defined by $\left(E + i\eta - \widehat{H}\right) G^r(x, x', E) = \delta(x - x')$, for $-\infty \leq x, x' \leq \infty$, with $\widehat{H}$ the Hamiltonian without the boundary terms.
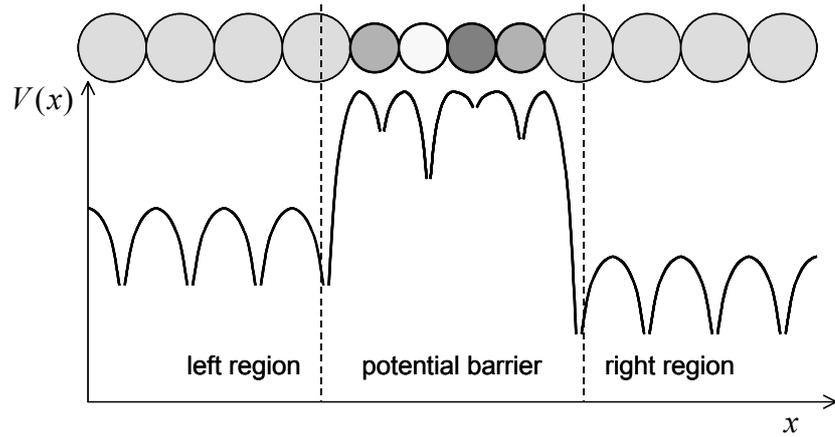
Figure 1.11: *Top*: atomic chain. The leads (left and right region) are periodic chains of identical atoms. The middle region contains different atoms and/or disorder. *Bottom*: schematic representation of the potential along the chain.

are represented by infinite dimensional matrices $\mathbf{A}$ with matrix elements $A_{i,j} = \langle \chi_i | \widehat{A} | \chi_j \rangle$ and the Schrödinger equation becomes

$$(E\mathbf{I} - \mathbf{H})\,\boldsymbol{\psi} = \mathbf{0}, \tag{1.92}$$

with $\mathbf{H}$ the Hamiltonian matrix.[26] To simplify the algebra we take just a single atomic orbital per atomic site. Moreover, we assume that the diagonal elements of $\mathbf{H}$ are $H_{i,i} = h_i$; its off-diagonal elements are $H_{i+1,i} = \beta_i$ and all elements $H_{j,i} = 0$ for $j > i + 1$. The other matrix elements are then set by demanding that the Hamiltonian matrix is Hermitian, i.e. $\mathbf{H}^\dagger = \mathbf{H}$.

$$\mathbf{H} = \begin{pmatrix} \ddots & \cdots & 0 & 0 & \\ \vdots & h_{i-1} & \beta_{i-1}^* & 0 & 0 \\ 0 & \beta_{i-1} & h_i & \beta_i^* & 0 \\ 0 & 0 & \beta_i & h_{i+1} & \vdots \\ & 0 & 0 & \cdots & \ddots \end{pmatrix}. \tag{1.93}$$

The model is called the **nearest neighbor** tight-binding approximation by physicists and among chemists it is known as the Hückel approximation. The approximation is not essential and the formalism that I will explain below can be made to work for any LCAO representation, but here I want to keep the expressions as simple as possible.[27]

We divide our system into three parts: a left lead, a scattering region, and a right lead. The left and right leads are perfect materials. They consist of identical atoms at equal distances a_L (for the left lead) and a_R (for the right lead). In other words, the leads have translational symmetry. Matrix elements in the leads must be identical, i.e. $h_i = h_{L/R}$ and $\beta_i = \beta_{L/R}$ for the left/right leads. Only in the scattering region do we have site dependent matrix elements. The basic idea is illustrated in Fig. 1.12.

---

[26]I want to keep things as simple as possible, so I am using an orthogal basis. If the atomic orbital basis is non-orthogonal, then introduce an overlap matrix $\mathbf{S}$, with matrix elements $S_{i,j} = \langle \chi_i | \chi_j \rangle$, and substitute $\mathbf{I}$ by $\mathbf{S}$.

[27]In one-dimensional tight-binding one can assume that the hopping parameters β_i are real, without loss of generality. However, in three dimensions or in magnetic fields this need no longer be true. Therefore, to be prepared for the more general case, I keep complex β_i's for the moment.
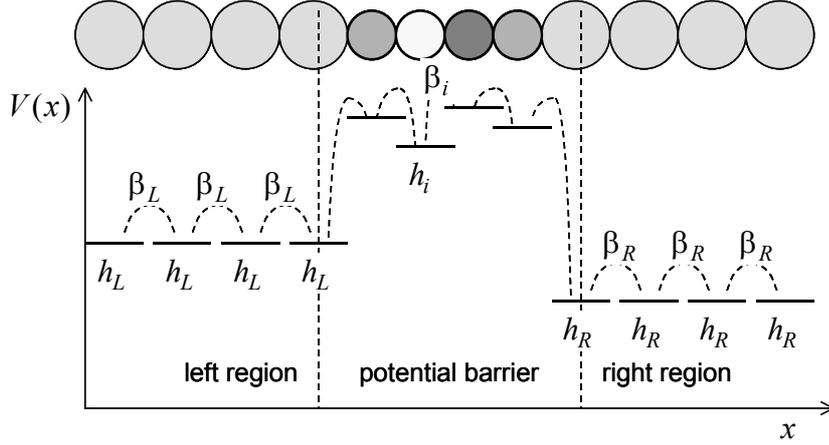
Figure 1.12: Nearest neighbor tight-binding model of an atomic chain. The periodic left and right regions are characterized by the on-site and hopping matrix elements $h_{L/R}$ and $\beta_{L/R}$. The scattering region has site dependent matrix elements h_i and β_i.

1.5.1   Ideal lead modes

Writing out Eq. 1.92 in its components gives

$$-\beta_{i-1}c_{i-1} + (E - h_i)c_i - \beta_i^* c_{i+1} = 0, \tag{1.94}$$

with i running from $-\infty$ to ∞. These equations have the same mathematical form as the discretized Schrödinger equation of Eq. 1.58 if we make the substitutions

$$\psi_i \to c_i; \quad -\frac{\hbar^2}{m\Delta^2} + V_i \to h_i; \quad -\frac{\hbar^2}{2m\Delta^2} \to \beta_i.$$

We are going to solve the scattering problem by mode matching. First we have to find the modes of the ideal leads. For sites i in the left lead the matrix elements are site independent and Eq. 1.94 becomes

$$-\beta_L c_{i-1} + (E - h_L)c_i - \beta_L^* c_{i+1} = 0. \tag{1.95}$$

A little pondering shows that this equation is mathematically the same as a discretized Schrödinger equation for a particle in a constant potential. Since "the same equations have the same solutions", the solutions must be (discretized) plane waves, i.e.

$$c_n = A e^{ik_L n a_L}, \tag{1.96}$$

where the distance between the atoms a_L is the "discretization step". The same holds for the right lead, replacing the subscript L by R.

One can give it a somewhat more mathematical flavor. The Bloch-Floquet theorem states that functions in consecutive cells of a periodic system are related by a constant amplitude/phase factor λ, i.e.[28]

$$\text{if } c_{i-1} = c \text{ then } c_i = \lambda c \text{ and } c_{i+1} = \lambda^2 c. \tag{1.97}$$

---

[28]In the physics literature this is known as the "Bloch theorem", and in the mathematics literature as the "Floquet theorem". Google a bit if you want to know the history.

Now is the time where we simplifying things a little bit by choosing β real,[29] so Eq. 1.95 becomes

$$-\beta + (E - h)\lambda - \beta\lambda^2 = 0 \Rightarrow \tag{1.98}$$

$$\lambda = \frac{E - h}{2\beta} \pm \left[\left(\frac{E - h}{2\beta}\right)^2 - 1\right]^{\frac{1}{2}},$$

where you can add the subscript L/R for the left and right leads. The roots λ can be given a more familiar form. For $\left|\frac{E-h}{2\beta}\right| \leq 1$ we define a wave number k by[30]

$$\cos(ka) = \frac{E - h}{2\beta}, \tag{1.99}$$

which leads to the simple form

$$\lambda_\pm = e^{\pm ika}. \tag{1.100}$$

$\lambda_\pm$ is called the **Bloch factor**. Using Eq. 1.97 recursively, i.e. $c_n = \lambda^n c_0$, then leads to Eq. 1.96, with $A = c_0$, as expected. It describes **propagating waves**, where λ_+ describes a wave propagating to the right, and λ_- a wave propagating to the left.

For $\left|\frac{E-h}{2\beta}\right| > 1$ one can define κ by

$$\cosh(\kappa a) = \left|\frac{E - h}{2\beta}\right|, \tag{1.101}$$

and obtain

$$\lambda_\pm = +e^{\mp\kappa a} \ \text{ if } \ \frac{E - h}{2\beta} > 1;$$

$$\lambda_\pm = -e^{\mp\kappa a} \ \text{ if } \ \frac{E - h}{2\beta} < -1. \tag{1.102}$$

Both these cases describe states that decay either to the right or to the left. These are called **evanescent states**. They are not acceptable as solutions to the one-dimensional Schrödinger equation because one cannot normalize them (not even in the wave packet sense). However, we will have a use for them later on in three-dimensional problems.

1.5.2 Mode matching

We have the modes of the ideal leads, so we can match them to the scattering region, where the matrix elements h_i and β_i in Eq. 1.94 are site dependent. We assume that the scattering region is localized in space, so i runs from 1 to N. The procedure we have used to solve the discretized Schrödinger problem in Sec. 1.4 can be copied with only some small modifications. Using Eq. 1.100, Eqs. 1.61 and 1.64 read

$$c_{-1} = A\lambda_{L,+}^{-1} + B\lambda_{L,-}^{-1} = A\lambda_{L,+}^{-1} + (c_0 - A)\lambda_{L,-}^{-1}; \tag{1.103}$$

$$c_{N+2} = c_{N+1}\lambda_{R,+}.$$

---

[29]which is always possible in one-dimensional systems, provided one has spin degeneracy and time-inversion symmetry, which we have here.

[30]You may recognize $E(k) = h + 2\beta \cos ka$ as the dispersion relation describing an electronic band in the nearest neighbor tight-binding model.

The convention is to write $c_{i+n} = \lambda^n c_i$, where n is an integer (positive or negative), and let the $\pm$ indicate waves propagating to the left and right, respectively.[31] These relations can be used to substitute Eq. 1.92 by

$$\left(E\mathbf{I} - \mathbf{H}'\right)\psi = \mathbf{q}, \tag{1.104}$$

similar to Eq. 1.66. ψ is a finite dimensional vector that contains the coefficients c_i in the scattering region plus those of the two boundaries, i.e. $i = 0, \dots, N+1$. $\mathbf{q}$ is the "source" vector of length $N + 2$, whose coefficients are zero, except the first one

$$q_0 = \beta_L A \left\{ \lambda_{L,-}^{-1} - \lambda_{L,+}^{-1} \right\}. \tag{1.105}$$

$\mathbf{H}'$ is a finite $(N+2) \times (N+2)$ Hamiltonian matrix. All its matrix elements are identical to that of the original Hamiltonian matrix, Eq. 1.93, except for the first and the last diagonal element, which are modified to

$$\begin{aligned} H'_{0,0} &= h_0 + \Sigma_L(E); \\ H'_{N+1,N+1} &= h_{N+1} + \Sigma_R(E), \end{aligned} \tag{1.106}$$

with

$$\begin{aligned} \Sigma_L(E) &= \beta_L \lambda_{L,-}^{-1}; \\ \Sigma_R(E) &= \beta_R \lambda_{R,+}. \end{aligned} \tag{1.107}$$

These self-energies contain all the information concerning the coupling of the scattering region to the leads. As before, they are complex and energy dependent through Eqs. 1.99 and 1.100.

We have substituted an infinite dimensional problem, Eq. 1.92, by a finite dimensional one, Eq. 1.104!! [32] Mathematically Eq. 1.104 is the same as Eq. 1.66. Again "the same equations have the same solutions", so according to Eq. 1.72, the transmission amplitude becomes

$$t = \sqrt{\frac{v_R}{v_L}} \frac{c_{N+1}}{A} \tag{1.108}$$

with, as in Eq. 1.77, the velocities given by

$$v_{L/R} = -\frac{2a_{L/R}}{\hbar} \operatorname{Im} \Sigma_{L/R}(E). \tag{1.109}$$

In addition, the Green function expressions given in Sec. 1.4.1 remain valid.

---

[31]In the one-dimensional case, one always has $\lambda_+ = 1/\lambda_-$. In the three-dimensional case, this relation does not necessarily hold. That's why it is important to keep separate track of the powers of λ and the $\pm$ indices.

[32]For those of you who have some background in this, you might suspect that the technique we are using here has something to do with a technique known as **partitioning**. This suspicion is appropriate. Partitioning is usually applied to operators and, in particular, to Green functions. I am applying it to wave functions here.

Chapter 2

Electron transport in three dimensions

In this chapter I want to upgrade the discussion of the previous chapter to problems in three dimensions. The tunnel junction of Fig. 1.2, for instance, is clearly not a strictly one-dimensional device. In Sec. 2.1 we will generalize Landauer's formula for the conductance to three dimensions. I will do this in steps of increasing complexity, i.e. start with plane waves and square barriers, as in the one-dimensional (1D) case, and end with atomistic three-dimensional (3D) devices. The techniques that are introduced should allow us to study transport at the nanoscale in any small system. In a tunnel junction the dimension of the device across the junction (determined by the thickness of the insulator layer) is usually much smaller that the dimensions parallel to the metal insulator interfaces. The modeling then becomes easier if one assumes the latter dimensions to be infinite and one uses translation symmetry along the interfaces. The latter requires doing a bit of solid state physics. Another "device" is shown in Fig. 2.8 below. It consists of a mono-atomic wire sandwiched between two electrodes. Such wires have been made of gold atoms, for instance. Clearly the wire is finite in all three dimensions and the electrodes are infinite. Our modeling must be able to cope both with finite and with infinite dimensions.

As in the previous chapter, Landauer's formula expresses the conductance in terms of the transmission of incoming electron waves. In three dimensions the transmission amplitudes constitute a matrix, the **transmission matrix**, whose calculation is our primary task. Practical algorithms are explained in Sec. 2.2. In particular, the mode matching techniques of Sec. 1.4 are generalized to 3D. Since in 3D the derivations of the expressions become quite lengthy and technical, I won't give all details. They can be found in the literature.

2.1 Landauer in 3D

2.1.1 Conductance of model interfaces

We start with a generalization of the square barrier potential of Fig. 1.3. A two-dimensional example is shown in Fig. 2.1. The potential $V(x, y)$ is separable; the barrier is in the x-direction and the potential is independent of y. The straightforward extension to three dimensions is a potential that is independent of y and z. Such a potential landscape is a simple model for a thin layer of an insulator sandwiched between two metals, i.e. a tunnel junction.

The Schrödinger equation

$$E\psi(\mathbf{r}) + \frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) - V(\mathbf{r})\psi(\mathbf{r}) = 0 \tag{2.1}$$
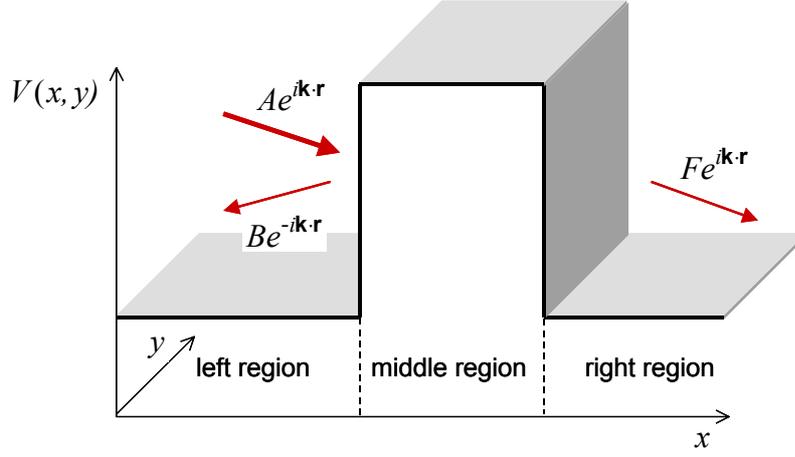
Figure 2.1: A square barrier in two dimensions. In the left and right regions the potential is constant, $V(x,y) = V_1$. In the middle region the potential is also constant, $V(x,y) = V_0$, where $V_0 > V_1$. The incoming, reflected and transmitted waves are given by $Ae^{i\mathbf{k}\cdot\mathbf{r}}$, $Be^{i\mathbf{k'}\cdot\mathbf{r}}$ and $Fe^{i\mathbf{k}\cdot\mathbf{r}}$, see Fig. 2.2.

is separable in Cartesian coordinates

$$\left[E + \left\{ \frac{\hbar^2}{2m}\frac{d^2}{dx^2} - V(x) \right\} + \frac{\hbar^2}{2m}\frac{d^2}{dy^2} + \frac{\hbar^2}{2m}\frac{d^2}{dz^2} \right] \psi(x,y,z) = 0,$$

and its solutions can be written as

$$\psi(x,y,z) = \phi(x)e^{ik_y y}e^{ik_z z} = \phi(x)e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}; \; \mathbf{k}_\parallel = \begin{pmatrix} 0 \\ k_y \\ k_z \end{pmatrix}, \tag{2.2}$$

where $e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}$ describes a free particle in the direction parallel to the barrier. The function $\phi(x)$ is the solution of the one-dimensional scattering problem

$$\left[E_x + \left\{ \frac{\hbar^2}{2m}\frac{d^2}{dx^2} - V(x) \right\} \right] \phi(x) = 0, \tag{2.3}$$

with the energy

$$E_x = E - \frac{\hbar^2 k_\parallel^2}{2m}.. \tag{2.4}$$

Note that $\mathbf{k}_\parallel$ is a good quantum number[1], which together with the energy E fixes the wave function. We say that $\mathbf{k}_\parallel$ defines a **mode** of the system.

A view of the scattering geometry in the $(k_x, \mathbf{k}_\parallel)$ plane is given in Fig. 2.2. In the left and right regions the wave function is

$$\psi_{\mathbf{k}_\parallel}(\mathbf{r}) = \begin{cases} A\left[e^{i\mathbf{k}\cdot\mathbf{r}} + r_{\mathbf{k}_\parallel}(E)e^{i\mathbf{k'}\cdot\mathbf{r}} \right]; \; \mathbf{r} \text{ in left region} \\ A\left[t_{\mathbf{k}_\parallel}(E)e^{i\mathbf{k}\cdot\mathbf{r}} \right]; \; \mathbf{r} \text{ in right region} \end{cases} \tag{2.5}$$

_____

[1] a set of two quantum numbers, actually.
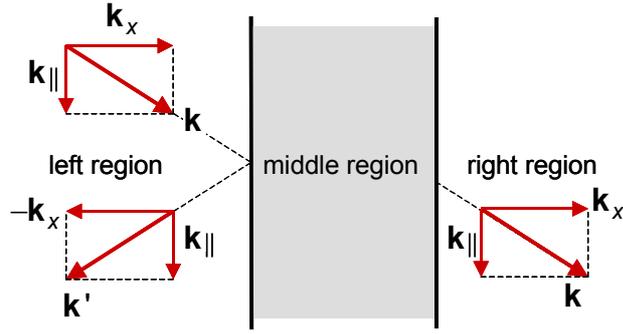
Figure 2.2: Scattering from a planar square barrier as viewed in the $(k_x, \mathbf{k}_\parallel)$ plane.

with $\mathbf{k} = (k_x, \mathbf{k}_\parallel)$ and $\mathbf{k}' = (-k_x, \mathbf{k}_\parallel)$. I leave you the job of finding the expressions for the transmission and reflection amplitudes $t_{\mathbf{k}_\parallel}(E)$ and $r_{\mathbf{k}_\parallel}(E)$. In three dimensions the probability current is a vector

$$\mathbf{J}(\mathbf{r}, t) = \frac{i\hbar}{2m} \left[\Psi(\mathbf{r}, t) \nabla \Psi^*(\mathbf{r}, t) - \Psi^*(\mathbf{r}, t) \nabla \Psi(\mathbf{r}, t) \right] \tag{2.6}$$

As in the one-dimensional case, for a stationary problem, $\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) e^{-\frac{i}{\hbar} Et}$, the current is constant. In three dimensions it is actually more appropriate to use the phrase "current density" for $\mathbf{J}$ (as in classical electrodynamics), but I keep on using the phrase "current" for short.

The key point

For the experiments we are interested in, **only J_x matters**, i.e. the x-component of the current. For devices like that shown in Fig. 2.1 (and all other devices that we will consider), one attaches macroscopically large electrodes to the left and right regions and applies a bias voltage between those electrodes. Only the current along the x-direction is then measured.[2,3] So we are interested in

$$J_x = \frac{i\hbar}{2m} \left[\psi(\mathbf{r}) \frac{\partial \psi^*(\mathbf{r})}{\partial x} - \psi^*(\mathbf{r}) \frac{\partial \psi(\mathbf{r})}{\partial x} \right]. \tag{2.7}$$

From Eq. 2.5 one can easily show that the transmitted current carried by one mode $\mathbf{k}_\parallel$ is given by

$$J_{T, \mathbf{k}_\parallel} = \rho \left| t_{\mathbf{k}_\parallel}(E) \right|^2 v_x \tag{2.8}$$

with the density $\rho = |A|^2$ and the velocity in the x-direction $v_x = \frac{\hbar k_x}{m}$. Deriving an expression for the conductance follows the same steps as in Sec. 1.2. Applying a small bias $\Delta V = -eU$

---

[2]For those of you who are experts in scattering phenomena, note that this is different from what you are used to in angle resolved three-dimensional scattering experiments in free space. There one is usually interested in all three components of the current. In addition, if the scatterer is localized in space, one applies a transformation to spherical coordinates.

[3]More complicated measurements are possible. In "multiprobe" experiments more than two electrodes are attached to the device. This is often done in combination with applying an external magnetic field, as in measuring the Hall effect. The Landauer formalism can be extended to include multiprobe measurements. This extension is due to Büttiker.

between left and right regions, the transmitted current is carried by all modes that have an energy $E_\mathbf{k}$ in the range from $E_F - \Delta V$ to E_F, see Fig. 1.7.

$$J_T = 2\rho \sum_{E_F - \Delta V < E_\mathbf{k} < E_F} \left| t_{\mathbf{k}_\parallel}(E) \right|^2 v_x, \tag{2.9}$$

where the factor 2 accounts for the degeneracy. We can use the same tricks as in Sec. 1.2. Using states that are normalized in a 3D, we have $\rho = \frac{1}{2L^3}$; compare to Eq. 1.32.[4] We write $\sum_{E_F - \Delta V < E_\mathbf{k} < E_F} = \sum_{\mathbf{k}_\parallel} \sum_{k_x}$, where one has to sum only over those states that have their energy in the indicated interval. Convert $\sum_{k_x}$ into a one-dimensional integral, use $v_x = \frac{1}{\hbar} \frac{dE_x}{dk_x}$, and assume that $t_{\mathbf{k}_\parallel}(E)$ is independent of the energy in the small energy range ΔV. The algebra is

$$
\begin{aligned}
J_T &= \frac{1}{L^3} \sum_{\mathbf{k}_\parallel} \sum_{k_x} \left| t_{\mathbf{k}_\parallel}(E) \right|^2 v_x = \frac{1}{L^3} \sum_{\mathbf{k}_\parallel} \frac{L}{\pi} \int \left| t_{\mathbf{k}_\parallel}(E) \right|^2 v_x dk_x \\
&= \frac{1}{L^2} \frac{1}{\pi} \sum_{\mathbf{k}_\parallel} \int \left| t_{\mathbf{k}_\parallel}(E) \right|^2 \frac{1}{\hbar} \frac{dE_x}{dk_x} dk_x = \frac{1}{L^2} \frac{1}{\pi\hbar} \sum_{\mathbf{k}_\parallel} \int_{E_F - \Delta V - \frac{\hbar^2 k_\parallel^2}{2m}}^{E_F - \frac{\hbar^2 k_\parallel^2}{2m}} \left| t_{\mathbf{k}_\parallel}(E_x + \frac{\hbar^2 k_\parallel^2}{2m}) \right|^2 dE_x \\
&= \frac{1}{L^2} \frac{\Delta V}{\pi\hbar} \sum_{\mathbf{k}_\parallel} \left| t_{\mathbf{k}_\parallel}(E_F) \right|^2.
\end{aligned}
\tag{2.10}
$$

The transmitted current is $I_T = \int \mathbf{J}_T \cdot d\mathbf{S} = J_T L^2$ and the conductance $\mathcal{G} = I_T / U$ is then given by

$$\mathcal{G} = \frac{e^2}{\pi\hbar} \sum_{\mathbf{k}_\parallel} \left| t_{\mathbf{k}_\parallel}(E_F) \right|^2 = \frac{e^2}{\pi\hbar} T(E_F). \tag{2.11}$$

The expression is the Landauer formula, see Eq. 1.29, with the total transmission T expressed as a sum over the transmissions of the individual modes $\mathbf{k}_\parallel$. One has to sum over the $\mathbf{k}_\parallel$ that contribute to the transmission at the Fermi energy E_F.

The Fermi surface

The Fermi surface is defined by the relation $E_\mathbf{k} = E_F$, the Fermi energy being a materials constant. It can be visualized as a surface in reciprocal space, i.e. in three-dimensional $\mathbf{k}$-space. For free electrons or electrons in a constant potential

$$E_\mathbf{k} = \frac{\hbar^2}{2m} \left(k_x^2 + k_y^2 + k_z^2 \right) = \frac{\hbar^2}{2m} \left(k_x^2 + k_\parallel^2 \right) = E_F.$$

the Fermi surface is the surface of a sphere, as shown in Fig. 2.3(a). The Fermi surface can help to visualize which modes contribute to the transmission. The latter can be enumerated by projecting the part of Fermi surface with $k_x > 0$ (a hemisphere in this case) onto the $\mathbf{k}_\parallel = (k_y, k_z)$ plane, as shown in Fig. 2.3(b). All the modes $\mathbf{k}_\parallel$ within this projection exist at the Fermi energy E_F and they contribute to the transmission in Eq. 2.11. The scattering geometry in real space can be deduced from Fig. 2.2. $\mathbf{k}_\parallel = (0,0)$[5] corresponds to a wave with normal incidence to the

---

[4]The function $\frac{1}{L^2} e^{i\mathbf{k}_\parallel \cdot \mathbf{r}}$ is normalized in a 2D box of size L, as you can easily check yourself. The function $\phi(x)$ is treated as in Eq. 1.32 and gets the normalization factor $\frac{1}{2L}$. The product gives the normalization factor of $\psi(x, y, z)$, see Eq. 2.2.

[5]called the Γ-point in solid state physics folklore.
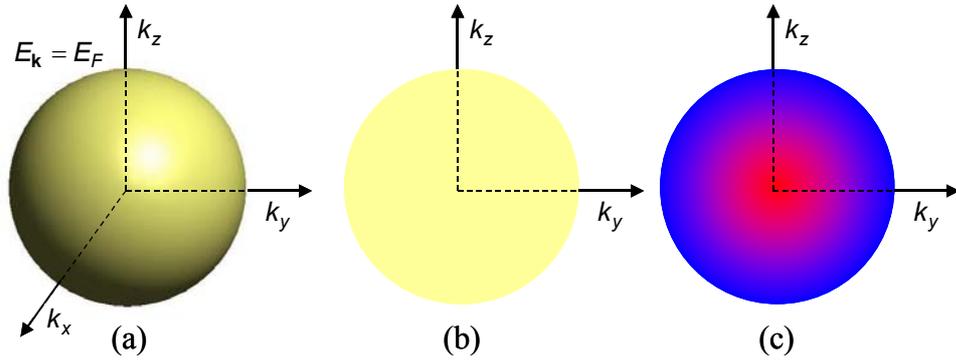
Figure 2.3: (a) The Fermi surface, as defined by $E_{\mathbf{k}} = E_F$, of electrons in a constant potential is a sphere. (b) The projection of the surface in the $\mathbf{k}_\parallel = (k_y, k_z)$ plane. The shaded area denotes all the $\mathbf{k}_\parallel$ modes that contribute to the transmission at E_F. (c) The transmission $\left|t_{\mathbf{k}_\parallel}\right|^2$ as function of $\mathbf{k}_\parallel$. Red indicates a high transmission and blue a low transmission. The highest transmission is for $\mathbf{k}_\parallel = (0,0)$. It decreases to 0 towards the edge of the circle.

barrier. The larger $k_\parallel = \left|\mathbf{k}_\parallel\right|$, the more glancing the incidence of the corresponding wave. At the edge of the circle in Fig. 2.3(b) one has $\frac{\hbar^2}{2m}k_\parallel^2 = E_F$, so $k_x = 0$, and the corresponding wave propagates parallel to the barrier.

For a simple square barrier one can easily calculate the transmission. If the Fermi energy is less than the barrier height, the transmission of a single mode is given by

$$\left|t_{\mathbf{k}_\parallel}(E_F)\right|^2 = T_{1D}(E_x) = T_{1D}(E_F - \frac{\hbar^2 k_\parallel^2}{2m}), \tag{2.12}$$

with T_{1D} given by Eqs. 1.47 and 1.42. The result is visualized in Fig. 2.3(c). Since T_{1D} is a monotonically increasing function of the energy, the maximal transmission is for $\mathbf{k}_\parallel = (0,0)$, i.e. for normal incidence. The transmission decreases monotonically with increasing $k_\parallel$, i.e. with the angle of incidence, until it is zero for parallel incidence. Such a simple Fermi surface and a simple transmission are typical for free electrons. For real materials both the Fermi surface and the transmission are much more complex. I will come back to in Sec. 2.1.4.

2.1.2 Conductance of model wires: the ballistic conductance

The way we have derived Eq. 2.11 means it can easily be generalized to any system whose Hamiltonian is separable in an x-term and a yz-term. For instance, the wave functions of the wire shown in Fig. 2.4 can be written as

$$\psi(x, \mathbf{r}_\perp) = \phi(x)\eta_{\mathbf{n}}(\mathbf{r}_\perp), \tag{2.13}$$

where $\mathbf{n}$ is a set of two quantum numbers labeling the modes; see Eq. 2.2. The conductance of this quantum wire can be expressed as the sum of the transmissions of the individual modes

$$\mathcal{G} = \frac{e^2}{\pi\hbar} \sum_{\mathbf{n}} |t_{\mathbf{n}}(E_F)|^2. \tag{2.14}$$

Suppose we are dealing with an ideal wire in which all the modes are fully transmitted.[6]

---

[6]which is the definition of an ideal wire.
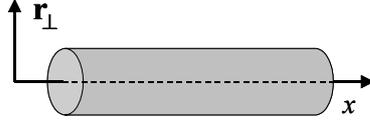
Figure 2.4: A simple (quantum) wire.

The transmission is maximal, i.e. $|t_\mathbf{n}(E_F)|^2 = 1$ for each mode $\mathbf{n}$, and the conductance is

$$\mathcal{G}_{\text{bal}} = \frac{e^2}{\pi\hbar} M(E_F), \tag{2.15}$$

where $M(E_F)$ is the number of modes at the Fermi energy, supported by the wire. $\mathcal{G}_{\text{bal}}$ is called the **ballistic conductance**. The number of modes at fixed energy in a wire of finite cross section is finite. So even for a perfect wire the conductance is finite!! If the cross section is macroscopically large, the number of modes is very large. Measuring the ballistic conductance experimentally is then impossible. A real wire always contains impurities and imperfections (lattice defects, grain boundaries, impurity atoms, etc.), and scattering at these defects dominates the conductance. However, thin and small wires can be made without any defects. Note that, since the number of modes is an integer, the **conductance** of Eq. 2.15 is **quantized**.

A very clear example of quantized conductance is presented by the experiment of van Wees *et al.* on the quantum conductance of a channel in a two-dimensional electron gas (2DEG). A 2DEG is formed at an interface between two well-chosen semiconductors. Putting electrodes on top of the 2DEG it is possible do define a one-dimensional channel (the "wire") by a suitable electrostatic potential profile, as illustrated in Fig. 2.5. The width of the channel determines the number of modes at the Fermi energy $M(E_F)$. Upon widening the channel, which is done by making the gate less repulsive for electrons, the number of modes increases and the conductance increases. However, it increases in steps of $\frac{e^2}{\pi\hbar}$. This is demonstrated by the experimental results, shown in Fig. 2.5.

2.1.3   Conductance of combined interfaces and wires

In the real world interfaces nor wires are infinite. This leads to Hamiltonians that are not separable in an x-term and a yz-term. A two-dimensional example is shown in Fig. 2.6, which represents a canyon through a 2D square barrier. It is a simple model for a finite wire connected to two electrodes (the left and right regions). An incoming wave $Ae^{i\mathbf{k}\cdot\mathbf{r}}$ with $\mathbf{k} = (k_x, \mathbf{k}_\parallel)$ can be scattered into any other wave $Ce^{i\mathbf{k}'\cdot\mathbf{r}}$ with $\mathbf{k}' = (k_x', \mathbf{k}_\parallel')$ provided the energy is conserved, i.e.

$$\frac{\hbar^2(k_x^2 + k_\parallel^2)}{2m} = E = \frac{\hbar^2(k_x'^2 + k_\parallel'^2)}{2m}. \tag{2.16}$$

The idea is shown in Fig. 2.7. The wave function of Eq. 2.5 is generalized to

$$\psi_{\mathbf{k}_\parallel}(\mathbf{r}) = \begin{cases} A\left[e^{i\mathbf{k}\cdot\mathbf{r}} + \sum_{\mathbf{k}_\parallel''} r_{\mathbf{k}_\parallel,\mathbf{k}_\parallel''}(E)e^{i\mathbf{k}''\cdot\mathbf{r}}\right] ; & \mathbf{r} \text{ in left region} \\ A\left[\sum_{\mathbf{k}_\parallel'} t_{\mathbf{k}_\parallel,\mathbf{k}_\parallel'}(E)e^{i\mathbf{k}'\cdot\mathbf{r}}\right] ; & \mathbf{r} \text{ in right region} \end{cases} \tag{2.17}$$

where $\mathbf{k}'$ labels the transmitted waves, i.e. $k_x' > 0$, and $\mathbf{k}''$ labels the reflected waves, i.e. $k_x'' < 0$, all at the same energy. The reflection and transmission amplitudes $r_{\mathbf{k}_\parallel,\mathbf{k}_\parallel''}$ and $t_{\mathbf{k}_\parallel,\mathbf{k}_\parallel'}$ indicate the probability amplitudes of scattering from a mode $\mathbf{k}_\parallel$ to a mode $\mathbf{k}_\parallel''$ or $\mathbf{k}_\parallel'$.
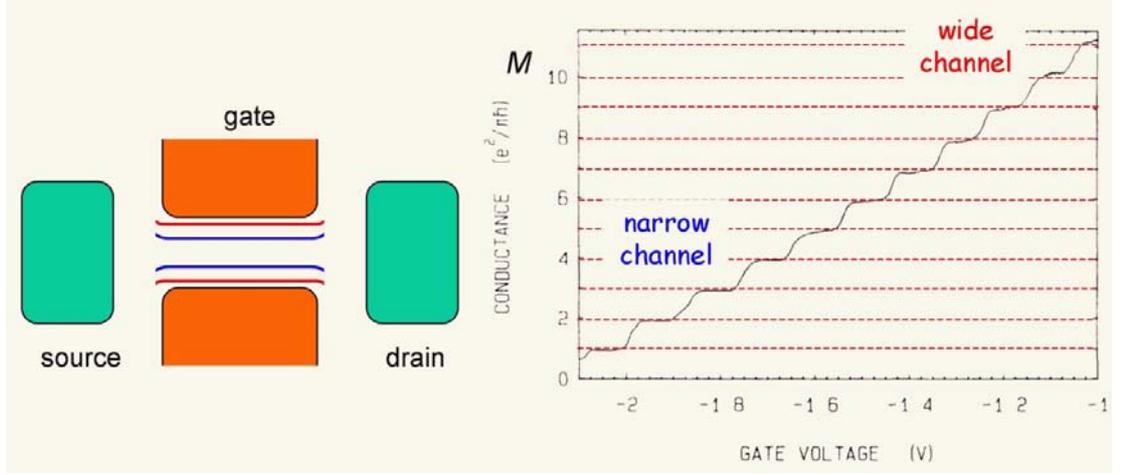
Figure 2.5: Quantized conductance of a ballistic waveguide *Left:* top view of the device. The current flows from the source to the drain electrode. The gate has a negative potential and repels the electrons towards the middle. This leaves an effective channel for the electrons. At a small gate potential the channel is wide (indicated in red), and at a large gate potential the channel is narrow (indicated in blue). *Right*: measured conductance as function of the gate potential. The conductance is quantized in units of $e^2/\pi\hbar$, as shown by the plateau's; M is the number of modes at E_F contributing to the conductance. See: B. J. van Wees *et al.*, Phys. Rev. Lett. **60**, 848 (1988).

The derivation of the Landauer formula follows the same steps as in the previous section. For instance, the current in the x-direction, carried by $\psi_{\mathbf{k}_\parallel}(\mathbf{r})$ is determined from Eq. 2.7

$$J_{T,\mathbf{k}_\parallel} = \rho \sum_{\mathbf{k}'_\parallel} \left| t_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel}(E) \right|^2 v'_x, \tag{2.18}$$

with $\rho = |A|^2$ and $v'_x = \frac{\hbar k'_x}{m}$; compare to Eq. 2.8. The generalization of Eq. 2.9 is

$$J_T = 2\rho \sum_{E_F-\Delta V < E_\mathbf{k} < E_F} \sum_{\mathbf{k}'_\parallel} \left| t_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel}(E) \right|^2 v'_x = 2\rho \sum_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel} \sum_{k'_x} \left| t_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel}(E) \right|^2 v'_x, \tag{2.19}$$

where the sum over the energies has been replaced by a sum over the states that have their energy in the right interval. Note that one does not have to sum over k_x. By fixing $\mathbf{k}_\parallel, \mathbf{k}'_\parallel$ and k'_x one has fixed k_x, because of Eq. 2.16. One can follow the steps of Eq. 2.10 and obtain an expression for the conductance which generalizes Eq. 2.11

$$\mathcal{G} = \frac{e^2}{\pi\hbar} \sum_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel} \left| t_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel}(E_F) \right|^2. \tag{2.20}$$

The transmission amplitudes $t_{\mathbf{k}_\parallel,\mathbf{k}'_\parallel}$ are collected in a **transmission matrix t**, which allows the conductance to be expressed in a compact form as

$$\mathcal{G} = \frac{e^2}{\pi\hbar} \mathrm{Tr}\left[\mathbf{t}^\dagger \mathbf{t} \right] = \frac{e^2}{\pi\hbar} T(E_F). \tag{2.21}$$

This is the **Landauer formula** for **non-uniform wires and interfaces**.
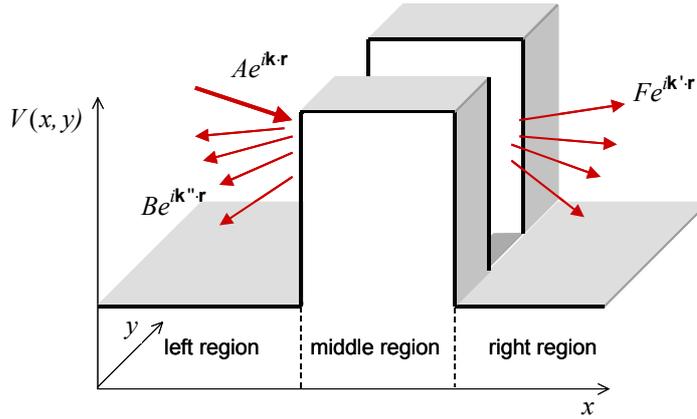
Figure 2.6: A simple model for a wire connected to two electrodes (the left and right regions) as a canyon through a square barrier. The incoming wave $Ae^{i\mathbf{k}\cdot\mathbf{r}}$ can be reflected to any wave $Be^{i\mathbf{k}''\cdot\mathbf{r}}$ or transmitted to any wave $Fe^{i\mathbf{k}'\cdot\mathbf{r}}$, provided the energy is conserved.
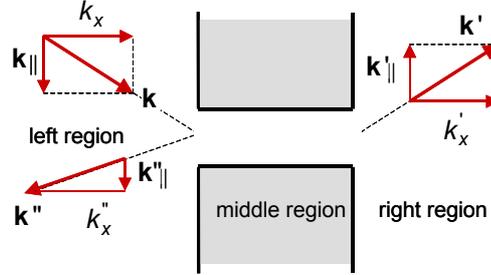
Figure 2.7: Scattering from a finite wire between two electrodes as viewed in the $\left(k_x, \mathbf{k}_\parallel\right)$ plane. Only one of the possible reflected $\left(k_x'', \mathbf{k}_\parallel''\right)$ waves and one of the transmitted $\left(k_x', \mathbf{k}_\parallel'\right)$ waves are indicated.

2.1.4   Conductance of atomistic wires and interfaces

We derived the Landauer formula, Eq. 2.21, using a model in which the incoming, reflected and transmitted waves in the asymtotic regions are plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$.[7] One might wonder how much of this derivation depends on having plane waves. Not too much, actually. We need to know what the waves in the asymtotic region look like, but they do not need to be plane waves. Consider the example shown in Fig. 2.8. It shows a mono-atomic wire of three atoms between two planar electrodes. Such wires have been made of gold and other metals (albeit with a less ideal structure).[8] The scattering region comprises the wire and a region around the wire in which the atomic potentials are different from bulk atomic potentials. Exactly how large that region is, depends somewhat on the system, but it will be clear that far into the electrodes, i.e. far into the left or right leads, the atoms are bulk metal atoms.[9]

---

[7]The regions outside the scattering region, i.e. the left and right regions are called the asymtotic regions.

[8]by a technique called "mechanical break junctions", for instance, see J. van Ruitenbeek.

[9]For metallic electrodes the potential at a few atomic layers from the surface is usually indistinguishable from the bulk potential, because screening effects in metals are large.
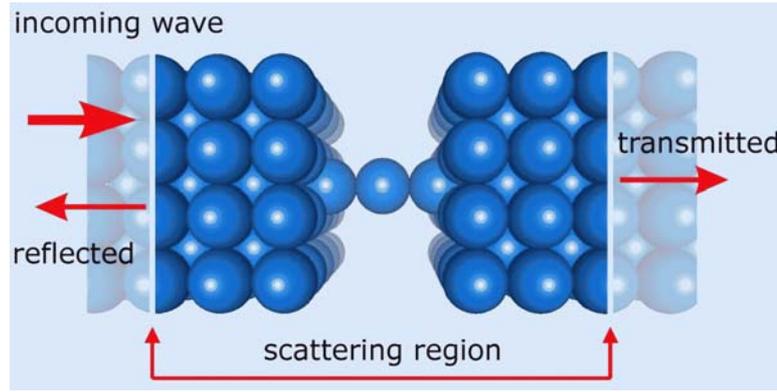
Figure 2.8: Schematic mono-atomic wire consisting of three atoms between two metal electrodes. Outside the scattering region one considers the electrodes as bulk metals, whose electronic states are represented by Bloch waves.

Bloch modes

Assuming that the **bulk leads** consist of **perfect crystalline material**, the Bloch-Floquet theorem tells us that the electronic states are represented by **Bloch waves**, which have the form

$$\phi_{\mathbf{k}n}(\mathbf{r}) = u_{\mathbf{k}n}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{2.22}$$

Here $u_{\mathbf{k}n}(\mathbf{r})$ is a periodic function in three dimensions, and the primitive unit corresponding to the three periods is called the unit cell. $\mathbf{k}$ is a wave vector within the reciprocal unit cell (called the first Brillouin zone) and n is the band index (a quantum number that distinguishes between states with the same wave vector). Proof of this can be found in any book on solid state physics. These Bloch waves replace the plane waves as modes in the scattering formalism.

The relation between energy and wave vector $E_{\mathbf{k}n}$ is called a **dispersion relation**. It can be obtained by solving the Schrödinger equation for electrons in the periodic crystal potential. As before, the **Fermi surface** is defined by the relation $E_{\mathbf{k}n} = E_F$, which gives a surface in reciprocal space, i.e. in $\mathbf{k}$-space, as in 2.3(a). Actually, it gives a surface for each n, called a **sheet** of the Fermi surface.

Fig. 2.9 gives a few examples of Fermi surfaces for different metals. The Fermi surface of some metals resembles that of free electrons; compare Figs. 2.9(a) and 2.3(a). These are called "simple metals"; they usually involve s- and/or p-electrons only. In other cases the Fermi surface can be quite complicated. It can consist of more than one sheet and has a shape that is far from free electron like (i.e. spherical), as is shown in 2.3(b). This is common among metals involving d-electrons.

The Landauer formula is still valid, however. The conservation of energy, Eq. 2.16, has to be replaced by the more general relation

$$E_{\mathbf{k}n} = E = E_{\mathbf{k}'n'}, \tag{2.23}$$

which basically states that one can scatter from a state $\mathbf{k}n$ to any state $\mathbf{k}'n'$, provided both points are on the same energy surface (the Fermi surface). In Eq. 2.17 one has to replace the plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}, e^{i\mathbf{k}'\cdot\mathbf{r}}, e^{i\mathbf{k}''\cdot\mathbf{r}}$ by Bloch waves $\phi_{\mathbf{k}n}(\mathbf{r}), \phi_{\mathbf{k}'n'}(\mathbf{r}), \phi_{\mathbf{k}''n''}(\mathbf{r})$ and the scattering amplitudes now also contain the band index, i.e. $r_{\mathbf{k}_{\parallel}n,\mathbf{k}_{\parallel}''n''}$ and $t_{\mathbf{k}_{\parallel}n,\mathbf{k}_{\parallel}'n'}$. Another vital ingredient needed
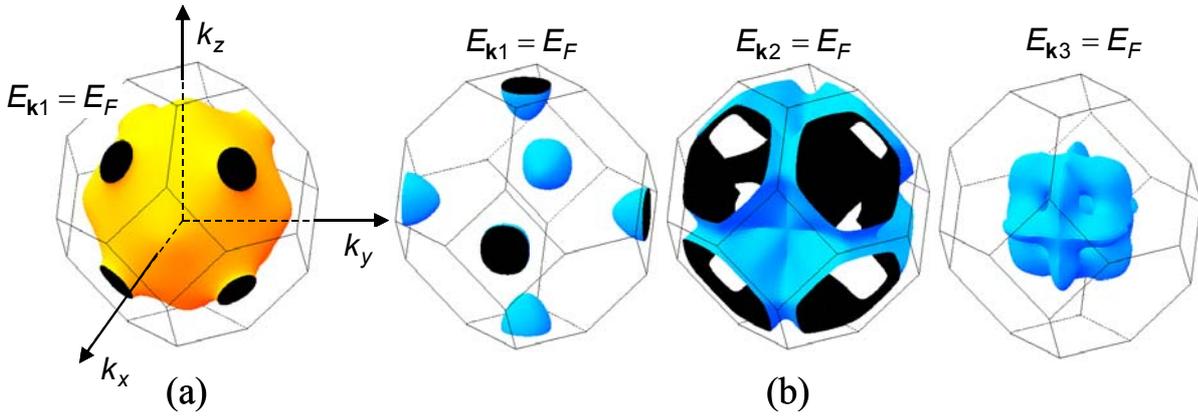
Figure 2.9: (a) The Fermi surface of Cu (copper). The wireframe indicates the first Brillouin zone (the reciprocal unit cell). The Fermi surface is almost spherical with a few holes punched through it. In Cu there is only one s-like band crossing the Fermi energy, which gives rise to one sheet. (b) The Fermi surface of fcc Co (Cobalt) for the minority spin electrons. Co is a ferromagnetic material. Only one s-like majority spin band crosses the Fermi energy, which gives rise to a Fermi surface for majority spin electrons that resembles that of Cu. Three d-like minority spin bands cross the Fermi energy, giving rise to three sheets of the Fermi surface that are far from spherical, as is shown here.

is an expression for the velocity v'_x. In solid state textbooks it is shown that the Bloch velocity is given by

$$v'_x = \frac{1}{\hbar} \frac{\partial E_{\mathbf{k}'n'}}{\partial k'_x}.$$ (2.24)

A little pondering shows that this is just the relation we need to complete the derivation of the Landauer formula.

$$\mathcal{G} = \frac{e^2}{\pi\hbar} \sum_{\mathbf{k}_\parallel n, \mathbf{k}'_\parallel n'} \left| t_{\mathbf{k}_\parallel n, \mathbf{k}'_\parallel n'}(E_F) \right|^2 = \frac{e^2}{\pi\hbar} \mathrm{Tr}\left[\mathbf{t}^\dagger \mathbf{t} \right].$$ (2.25)

An example: the conductance of an interface between copper and cobalt

Does it matter whether one takes real metals with Bloch modes as electrodes, or jellium metals, where the modes are plane waves? As an example that it does matter I will discuss the conductance of a simple interface between the metals copper (Cu) and cobalt (Co). Bulk Cu has a fcc (face centered cubic) close packed structure. Its (111) surface is close packed and it forms the crystal termination with the lowest surface energy. Bulk Co has the hcp (hexagonally close packed) structure. However if one deposits a thin layer of Co on Cu(111), then Co assumes a fcc structure with the lattice parameter of Cu. So the (111) interface between Cu and Co has a simple structure. So far we have considered spin degenerate systems, i.e. spin-up electrons are described by the same wave functions as spin-down electrons. This hold for Cu, which is a non-magnetic material. It does not hold for Co, which is a magnetic material, and both spins have different wave functions. In addition the number of spin-up electrons is not equal to the number of spin-down electrons (which is why the material is magnetic). The obvious names majority spin and minority spin are used.

The Landauer formula can be generalized to include the spin states explicitly in a straight-forward manner

$$\mathcal{G} = \frac{e^2}{2\pi\hbar} \sum_{\mathbf{k}_\| n\sigma, \mathbf{k}_\|' n'\sigma'} \left| t_{\mathbf{k}_\| n\sigma, \mathbf{k}_\|' n'\sigma'}(E_F) \right|^2, \tag{2.26}$$

where $\sigma, \sigma' = -\frac{1}{2}, \frac{1}{2}$. In absence of spin-orbit coupling the scattering potential does not flip the spin and the transmission matrix is diagonal in the spin $t_{\mathbf{k}_\| n\sigma, \mathbf{k}_\|' n'\sigma'} = t_{\mathbf{k}_\| n\sigma, \mathbf{k}_\|' n'\sigma} \delta_{\sigma\sigma'}$. This is accurate for (magnetic) 3d transition metals and lighter elements.

One can use the same procedure as in Fig. 2.3 to visualize the transmission. The Fermi surfaces of Cu and of the minority spin of Co are given in Fig. 2.9. The Fermi surface of the majority spin of Co resembles that of Cu. One has to project these surfaces onto the $\mathbf{k}_\|$ plane, i.e. along the transport direction, which is the (111) direction in this case. In Fig. 2.9(a) (111) is the direction pointing from the origin to the center of a hexagon. The projection of the Fermi surface of Cu is shown in Fig. 2.10(a). The projected Fermi surface for the majority spin of Co is shown in Fig. 2.10(b). As stated, it is quite similar to that of Cu. The Fermi surface for the minority spin of Co is much more complicated, as shown in Fig. 2.9 (b). Projecting all three sheets in the (111) direction on top of one another gives the result shown in Fig. 2.10(c).

For a perfect crystalline interface $\mathbf{k}_\|$ is a good quantum number and the transmission matrix is diagonal in $\mathbf{k}_\|$ (as well as in σ) $t_{\mathbf{k}_\| n\sigma, \mathbf{k}_\|' n'\sigma'} = t_{\mathbf{k}_\| n\sigma, \mathbf{k}_\| n'\sigma} \delta_{\mathbf{k}_\| \mathbf{k}_\|'} \delta_{\sigma\sigma'}$. For the majority spin $\left| t_{\mathbf{k}_\| 1\sigma, \mathbf{k}_\| 1\sigma} \right|^2$ describes the transmission from the state $\mathbf{k}_\| 1\sigma$ on the projected Fermi surface of Cu (indicated by a black dot in Fig. 2.10(a)) to a state $\mathbf{k}_\| 1\sigma$ on the projected majority spin Fermi surface of Co (indicated by a black dot in Fig. 2.10(b)).[10] The calculated transmission $\left| t_{\mathbf{k}_\| 1\sigma, \mathbf{k}_\| 1\sigma} \right|^2$ is shown in Fig. 2.10(d) in the $\mathbf{k}_\|$ plane. The transmission is a number between 0 and 1. Obviously the state $\mathbf{k}_\| 1\sigma$ has to exist on the Cu Fermi surface; otherwise one has no electrons to start with. Around $\mathbf{k}_\| = (0,0)$ the Cu Fermi surface has a "hole" where there are no states, which is indicated in white in 2.10(a) and (d). Then the state $\mathbf{k}_\| 1\sigma$ has to exist on the Co Fermi surface; otherwise there is no state to transport the electrons to and the transmission is 0. These are the blue area's in Fig. 2.10(d). The remaining area is almost entirely red, which means that there the transmission is close to 1.

Fig. 2.10(e) shows the calculated transmission $\sum_{n'=1}^{3} \left| t_{\mathbf{k}_\| 1\sigma, \mathbf{k}_\| n'\sigma} \right|^2$ between the minority spin modes of Cu and Co. As one can observe, the transmission pattern is much more complicated than in the majority spin case with a fine-grained variation of the transmission between 0 and 1. Now compare Figs. 2.10(d) and (e) to Fig. 2.3 (c). It represents the difference between using real metals and using jellium.[11]

Integrating over the area shown in Fig. 2.10(d) gives the majority spin conductance; the calculated number is $\mathcal{G}_{\mathrm{maj}} = \frac{e^2}{2\pi\hbar} 0.73$. Integrating over the area shown in Fig. 2.10(e) gives the minority spin conductance $\mathcal{G}_{\mathrm{min}} = \frac{e^2}{2\pi\hbar} 0.66$. The difference in the conductance between the majority and the minority spin modes is 10%. One can enlarge this difference by making a multilayer of alternating Cu and Co layers since such a multilayer then contains many Cu/Co interfaces. The majority/minority difference in the conductance essentially gives the so-called GMR (giant magneto-resistance) effect that is measured in such multilayers.[12]

---

[10] $n = 1$ and $n' = 1$, since both Fermi surfaces have only one sheet.

[11] Jellium also does not give magnetism at these electronic densities, but I don't want to make a list of what's wrong with jellium.

[12] The GMR effect is used to make sensitive sensors for magnetic fields, which are used in the heads of magnetic hard drives. It contributed to the rapid development in hard disks to the $> 10^2$ Gb disks we have today.
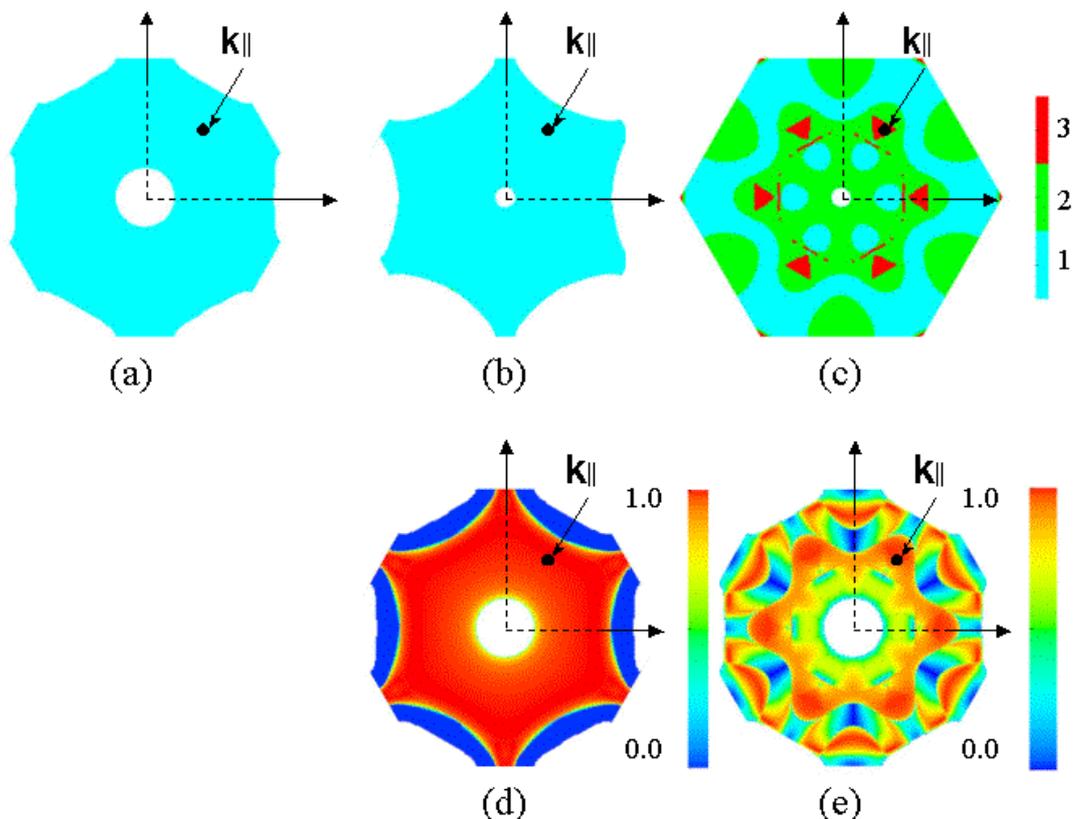
Figure 2.10: (a) Fermi surface of Cu projected in the (111) direction on the $\mathbf{k}_\parallel$ plane. (b) Fermi surface of the majority spin of Co projected on the same plane. (c) The three sheets of the Fermi surface of the minority spin of Co projected in the same plane. Blue, green and red indicate area's where 1,2 and 3 states are projected onto the same $\mathbf{k}_\parallel$ point. (d) The transmission across a (111) interface between the Cu and the majority spin in Co, as a function of $\mathbf{k}_\parallel$. The color scale indicates the transmission from 0 (blue) to 1 (red). (e) As (d), but for the transmission between Cu and the minority spin of Co. The white area's indicate where there are no states in Cu.

2.2   Calculating the transmission

As will be clear by now, the key quantity for calculating the conductance is the transmission matrix $\mathbf{t}$. This is quite a bit more complicated than in the one-dimensional case, but in principle all the techniques discussed in Secs. 1.4 and 1.5 can be extended to calculate transmission matrices in three dimensions for systems containing real atoms. Since the tight-binding example given in Sec. 1.5 is chemically the most intuitive, I use this and focus upon the mode matching technique.

The principal idea is to **divide** your system **into layers** of atoms, normal to the transport direction. One chooses the layers to be sufficiently thick such that the Hamiltonian matrix only contains matrix elements that either couple atoms within one layer, or atoms that are in nearest
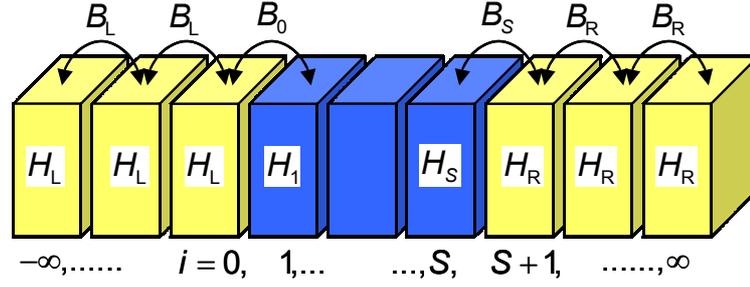
Figure 2.11: Hamiltonian of a tunnel junction divided into layers. The transport direction is along the horizontal. The left (L) and right (R) leads are ideal periodic wires containing the layers $i = -\infty, \dots, 0$ and $i = S+1, \dots, \infty$, respectively. The layers $i = 1, \dots, S$ consitute the scattering region.

neighbor layers.[13] The Hamiltonian matrix of Eq. 1.93 then becomes

$$\mathbf{H} = \begin{pmatrix} \ddots & \cdots & 0 & 0 & \\ \vdots & \mathbf{H}_{i-1} & \mathbf{B}_{i-1}^{\dagger} & 0 & 0 \\ 0 & \mathbf{B}_{i-1} & \mathbf{H}_i & \mathbf{B}_i^{\dagger} & 0 \\ 0 & 0 & \mathbf{B}_i & \mathbf{H}_{i+1} & \vdots \\ & 0 & 0 & \cdots & \ddots \end{pmatrix}. \tag{2.27}$$

A schematic representation of the structure of the Hamiltonian is given Fig. 2.11. The matrices $\mathbf{H}_i$ contain the interactions between atoms within the layer i, with $i = -\infty, \dots, \infty$, whereas the matrices $\mathbf{B}_i$ describe the coupling between the layers i and $i+1$. Both of these are $N \times N$ matrices, where N is the total number of atomic orbitals for all atoms in a layer. The scattering region is localized in the layers $i = 1, \dots, S$.[14]

This representation is valid both for systems with a finite cross section (i.e. wires), and for infinite layered systems that are periodic along the interfaces (i.e. along the layers). In the last case N refers to the number of atoms within the unit cell. $\mathbf{k}_{\parallel}$ is then a good quantum number and the matrices $\mathbf{H}_i(\mathbf{k}_{\parallel})$ depend on this quantum number. I won't discuss the exact expressions; they can be found in the literature. In fact, in order to simplify the notation I will omit the quantum number $\mathbf{k}_{\parallel}$ from now on. I will use the phrase "wire" to indicate both systems with a finite cross section and systems with infinite periodic interfaces.

The wave function of Eq. 1.91 is generalized to

$$\boldsymbol{\psi} = \begin{pmatrix} \vdots \\ \mathbf{c}_{i-1} \\ \mathbf{c}_i \\ \mathbf{c}_{i+1} \\ \vdots \end{pmatrix}, \tag{2.28}$$

---

[13]The formalism can be extended to include interactions of a longer range, but the expressions become a bit messy. So far there has been no need for this extension.

[14]Note that compared to Secs. 1.4 and 1.5 I have a slight change in notation here. Now N is the dimension of the basis *within* one layer, and S is the *number* of layers in the scattering region. Sorry about that; it's what you get if you merge texts that have a different origin.

where $\mathbf{c}_i$ is the N-vector of coefficients of the atomic orbitals of all atoms in a layer. As in Sec. 1.5 we divide our system into three parts, with $i = -\infty, \dots, 0$ corresponding to the left lead (L), $i = 1, \dots, S$ to the scattering region (S) and $i = S+1, \dots, \infty$ to the right lead (R). Using the mode matching technique the scattering problem is solved in two steps. In the first step the Bloch modes of the leads are calculated and in the second step these are then matched to the scattering region.

2.2.1   Ideal lead modes

The leads are assumed to be ideal wires characterized by a periodic potential. It is then appropriate to identify a layer with a translational period along the wire. By construction, the Hamiltonian matrix then is the same for each layer in the leads, i.e. $\mathbf{H}_i = \mathbf{H}_{L/R}$ and $\mathbf{B}_i = \mathbf{B}_{L/R}$ for the left/right leads, see Fig. 2.11. Eq. 1.95 is generalized to

$$-\mathbf{B}_{L/R}\mathbf{c}_{i-1} + (E\mathbf{1} - \mathbf{H}_{L/R})\mathbf{c}_i - \mathbf{B}_{L/R}^\dagger\mathbf{c}_{i+1} = 0, \tag{2.29}$$

where $\mathbf{1}$ is the $N \times N$ identity matrix.[15] We make the same ansatz as in Eq. 1.97, namely that the coefficients in successive layers are connected by a Bloch factor λ

$$\mathbf{c}_{i-1} \equiv \mathbf{c}; \ \mathbf{c}_i = \lambda\mathbf{c}; \ \mathbf{c}_{i+1} = \lambda\mathbf{c}_i = \lambda^2\mathbf{c} \tag{2.30}$$

to get the equation

$$-\mathbf{B}\mathbf{c} + \lambda(E\mathbf{1} - \mathbf{H})\mathbf{c} - \lambda^2\mathbf{B}^\dagger\mathbf{c} = 0, \tag{2.31}$$

where the subscripts L/R have been omitted to simplify the notation. Remember that we work at a fixed energy E, so Eq. 2.31 is a quadratic eigenvalue equation in λ of dimension N. There are standard tricks to solve such equations. For instance, by defining $\mathbf{d} = \lambda\mathbf{c}$ one can convert it into

$$\left[\begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{B} & E\mathbf{1} - \mathbf{H} \end{pmatrix} - \lambda \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^\dagger \end{pmatrix} \right] \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix} = 0. \tag{2.32}$$

This is a linear (generalized) eigenvalue problem of dimension $2N$, which can be solved using standard numerical routines.

It can be shown that this equation indeed generally has $2N$ solutions, which can be divided into N right-going modes and N left-going modes, labelled by a "+" and a "−" subscript as in Eqs. 1.100 and 1.102. Right-going modes are either evanescent waves that are decaying to the right, or waves of constant amplitude that are propagating to the right, whereas left-going modes are decaying or propagating to the left. Figs. 2.12 and 2.13 give you a simple idea.

In contrast to the 1D case, we find in three dimensions at a fixed energy in general **both evanescent and propagating modes**. We denote the eigenvalues and eigenvectors of Eq. 2.31 by[16]

$$\lambda_{\pm,n}; \ \mathbf{u}_{\pm,n}; \ n = 1, \dots, N. \tag{2.33}$$

Together these states form a complete basis set. In the following we assume that the vectors $\mathbf{u}_{\pm,n}$ are normalized. However, note that in general they are **not orthogonal**.[17] One can easily

---

[15] Again, for non-orthogonal basis sets one can substitute $\mathbf{1}$ by $\mathbf{S}$, the overlap matrix.

[16] Again I remind you that I have omitted the $\mathbf{k}_\parallel$ quantum number. So if you see the index n to denote a mode in the following, you can substitute it by $\mathbf{k}_\parallel n$ or by $\mathbf{k}_\parallel n\sigma$ to denote the full mode.

[17] This follows from the quadratic eigenvalue problem, Eq. 2.31, or the equivalent generalized linear one, Eq. 2.32.
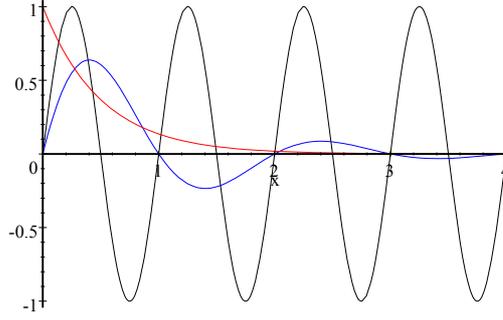
Figure 2.12: Right-going, i.e. "+", modes. Black curve: propagating mode, $|\lambda_+| = 1$. Blue and red curves: examples of evanescent modes, $|\lambda_+| < 1$.
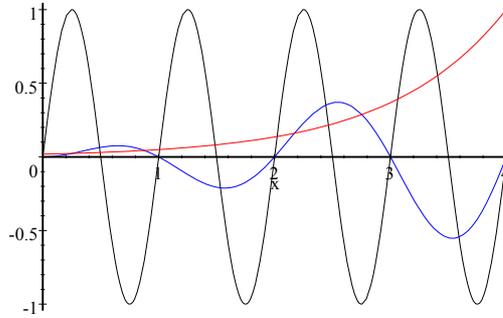
Figure 2.13: Left-going, i.e. "−", modes. Black curve: propagating mode, $|\lambda_-| = 1$. Blue and red curves: examples of evanescent modes, $|\lambda_-| > 1$.

distinguish right- from left-going evanescent modes on the basis of their eigenvalues. Right-going evanescent modes, which decay to the right, have $|\lambda_{+,n}| < 1$ and left-going evanescent modes, which decay to the right, have $|\lambda_{-,n}| > 1$, see Eq. 2.30 and Figs. 2.12 and 2.13. Propagating modes per definition have $|\lambda_{\pm,n}| = 1$, so here one has to determine the Bloch velocity in the propagation direction and use its sign to distinguish right from left propagation. One can show that for a tight-binding form of the Hamiltonian, the general expression for the Bloch velocities of Eq. 2.24 becomes

$$v_{\pm,n} = -\frac{2a}{\hbar} \operatorname{Im} \left[\lambda_{\pm,n} \mathbf{u}_{\pm,n}^\dagger \mathbf{B}^\dagger \mathbf{u}_{\pm,n} \right], \tag{2.34}$$

where a is the thickness of the layer. The derivation of this expression is rather technical and can be found in the literature. In addition one can show that the Bloch velocity is non-zero only for propagating modes, i.e. propagating states have a Bloch velocity equal to zero.[18]

Since the eigenvectors are non-orthogonal, it is convenient to define dual vectors $\widetilde{\mathbf{u}}_{\pm,n}$ by

$$\widetilde{\mathbf{u}}_{\pm,n}^\dagger \mathbf{u}_{\pm,m} = \delta_{n,m}; \ \mathbf{u}_{\pm,n}^\dagger \widetilde{\mathbf{u}}_{\pm,m} = \delta_{n,m}. \tag{2.35}$$

---

[18]Which is what you might expect. Evanescent modes contribute a zero current, i.e. no particles, energy, or whatever is transported by evanescent modes.

Any wave function in the leads can be expressed as a linear combination of the lead modes. This can be done in a very compact way by defining the two $N \times N$ **Bloch matrices** for right- and left-going modes

$$\mathbf{F}_{\pm} = \sum_{n=1}^{N} \lambda_{\pm,n} \mathbf{u}_{\pm,n} \widetilde{\mathbf{u}}_{\pm,n}^{\dagger}. \tag{2.36}$$

These are the generalizations of the Bloch factors of the one-dimensional case, see Eqs. 1.100 and 1.102. Note that one can easily construct powers of Bloch matrices

$$\mathbf{F}_{\pm}^{i} = \sum_{n=1}^{N} \lambda_{\pm,n}^{i} \mathbf{u}_{\pm,n} \widetilde{\mathbf{u}}_{\pm,n}^{\dagger}. \tag{2.37}$$

This is valid for any integer i, due to Eq. 2.35. A general solution in the leads can now be expressed as a recursion relation

$$\mathbf{c}_{i} = \mathbf{c}_{+,i} + \mathbf{c}_{-,i} = \mathbf{F}_{+}^{i-j} \mathbf{c}_{+,j} + \mathbf{F}_{-}^{i-j} \mathbf{c}_{-,j}. \tag{2.38}$$

In a scattering problem one usually fixes the coefficients in one layer using boundary conditions. By using Eq. 2.38 one can then determine the solution in all the layers of the leads.

2.2.2   Mode matching

One can use these recursion relations to set up a set of equations that properly match the leads to the scattering region. The scattering region is defined by the layers $i = 1, \dots, S$, see Fig. 2.11. Immediately left of the scattering regions one has the recursion relation

$$\mathbf{c}_{-1} = \mathbf{F}_{L,+}^{-1} \mathbf{c}_{+,0} + \mathbf{F}_{L,-}^{-1} \mathbf{c}_{-,0}.$$

Writing $\mathbf{c}_{-,0} = \mathbf{c}_0 - \mathbf{c}_{+,0}$ one gets

$$\mathbf{c}_{-1} = \left[\mathbf{F}_{L,+}^{-1} - \mathbf{F}_{L,-}^{-1} \right] \mathbf{c}_{+,0} + \mathbf{F}_{L,-}^{-1} \mathbf{c}_0. \tag{2.39}$$

The scattering boundary condition is introduced in the following way. The vector $\mathbf{c}_{+,0}$ is treated as the source, i.e. as the incoming wave from the left lead, e.g. a specific (propagating) mode of the left lead.

$$\mathbf{c}_{+,0} = \mathbf{u}_{L,+,m}. \tag{2.40}$$

Immediately right of the scattering region one has

$$\mathbf{c}_{S+2} = \mathbf{F}_{R,+} \mathbf{c}_{+,S+1} + \mathbf{F}_{R,-} \mathbf{c}_{-,S+1}.$$

Here the scattering boundary condition states that we have no incoming wave from the right lead, i.e. $\mathbf{c}_{-,S+1} = 0$, so

$$\mathbf{c}_{S+2} = \mathbf{F}_{R,+} \mathbf{c}_{S+1}. \tag{2.41}$$

The relations Eqs. 2.39 and 2.41 can now be used to "simplify" the tight-binding equations

$$-\mathbf{B}_i \mathbf{c}_{i-1} + (E\mathbf{1} - \mathbf{H}_i)\mathbf{c}_i - \mathbf{B}_i^{\dagger} \mathbf{c}_{i+1} = 0. \tag{2.42}$$

Ordinarily i runs from $-\infty$ to ∞, but for $i = 0$ we can write

$$
\begin{aligned}
-\mathbf{B}_L\mathbf{c}_{-1} + (E\mathbf{1} - \mathbf{H}_L)\mathbf{c}_0 - \mathbf{B}_1^\dagger\mathbf{c}_1 &= 0 \Leftrightarrow \\
(E\mathbf{1} - \mathbf{H}_L - \mathbf{B}_L\mathbf{F}_{L,-}^{-1})\mathbf{c}_0 - \mathbf{B}_1^\dagger\mathbf{c}_1 &= \mathbf{B}_L\left[\mathbf{F}_{L,+}^{-1} - \mathbf{F}_{L,-}^{-1}\right]\mathbf{u}_{L,+,m}.
\end{aligned}
\tag{2.43}
$$

Likewise one can write for $i = S + 1$

$$
\begin{aligned}
-\mathbf{B}_S\mathbf{c}_S + (E\mathbf{1} - \mathbf{H}_R)\mathbf{c}_{S+1} - \mathbf{B}_R^\dagger\mathbf{c}_{S+2} &= 0 \Leftrightarrow \\
-\mathbf{B}_S\mathbf{c}_S + (E\mathbf{1} - \mathbf{H}_R - \mathbf{B}_R^\dagger\mathbf{F}_{R,+})\mathbf{c}_{S+1} &= 0.
\end{aligned}
\tag{2.44}
$$

Eq. 2.42, for $i = 1, \dots, S$, and Eqs. 2.43 and 2.44 can be collected into

$$
(E\mathbf{1} - \mathbf{H}')\boldsymbol{\psi} = \mathbf{Q}_{L,+,m},
\tag{2.45}
$$

with

$$
\mathbf{H}' = \begin{pmatrix}
\mathbf{H}_L + \Sigma_L(E) & \mathbf{B}_1^\dagger & 0 & 0 & 0 \\
\mathbf{B}_1 & \ddots & \mathbf{B}_{i-1}^\dagger & 0 & 0 \\
0 & \mathbf{B}_{i-1} & \mathbf{H}_i & \mathbf{B}_i^\dagger & 0 \\
0 & 0 & \mathbf{B}_i & \ddots & \mathbf{B}_S^\dagger \\
0 & 0 & 0 & \mathbf{B}_S & \mathbf{H}_R + \Sigma_R(E)
\end{pmatrix},
\tag{2.46}
$$

and

$$
\boldsymbol{\psi} = \begin{pmatrix} \mathbf{c}_0 \\ \vdots \\ \mathbf{c}_i \\ \vdots \\ \mathbf{c}_{S+1} \end{pmatrix} ; \quad
\mathbf{Q}_{L,+,m} = \begin{pmatrix} \mathbf{B}_L\left[\mathbf{F}_{L,+}^{-1} - \mathbf{F}_{L,-}^{-1}\right]\mathbf{u}_{L,+,m} \\ \vdots \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}.
\tag{2.47}
$$

The quantities

$$
\begin{aligned}
\Sigma_L(E) &= \mathbf{B}_L\mathbf{F}_{L,-}^{-1}; \\
\Sigma_R(E) &= \mathbf{B}_R^\dagger\mathbf{F}_{R,+},
\end{aligned}
\tag{2.48}
$$

are called the **self-energies** of the left and right leads, respectively, just as in Eq. 1.107. They contain all the information about the coupling of the scattering region to the leads, as well as the information about the scattering boundary conditions. Note that the self-energies depend upon the energy E, since are expressed in the **Bloch matrices** and thus in the lead modes and the latter have been determined at a fixed energy E. Moreover, the self-energies are non-Hermitian, which makes the Hamiltonian $\mathbf{H}'$ **non-Hermitian**.[19] An important observation is that Eq. 2.45 has a **finite dimension**![20]

Eq. 2.45 represents a set of linear equations of dimension $(S + 2) \times N$. These can be solved using the common techniques, e.g. Gaussian elimination (or LU decomposition, as it is also

---

[19]This is logical if you know your quantum mechanics. A particle in the scattering region has a finite probability to leak into the leads, i.e. to disappear from the scattering region. In other words, the particle has a finite lifetime. This can only be expressed by non-Hermitian "Hamiltonians", as Hermitian ones would only give real energies and time factors $e^{-\frac{i}{\hbar}Et}$ yielding constant probabilities, i.e. infinite lifetimes.

[20]If you suspect that the "partitioning technique" is behind this, you are right.

known), followed by back substitution. Since the matrix $\mathbf{H}'$ has a block tridiagonal form (the blocks being of dimension N), one can make use of this special form to make the Gaussian elimination algorithm efficient. The details can be found in the literature. **Transmission matrix elements** are obtained by expanding the wave function in the right lead into modes[21]

$$\mathbf{c}_{S+1} = \sum_{n-1}^{N} \mathbf{u}_{R,+,n} t'_{n,m} \ \Rightarrow \ t'_{n,m} = \widetilde{\mathbf{u}}^{\dagger}_{R,+,n} \mathbf{c}_{S+1} \tag{2.49}$$

and normalize them with the velocities to ensure a unitary scattering matrix, see Eqs. 1.53 and 1.72.

$$t_{n,m} = \sqrt{\frac{v_{L,+,m}}{v_{R,+,n}}} \widetilde{\mathbf{u}}^{\dagger}_{R,+,n} \mathbf{c}_{S+1} \tag{2.50}$$

The velocities are calculated from Eq. 2.34.

By letting m in Eqs. 2.40 and 2.47 run over all (propagating) modes $\mathbf{u}_{L,+,m}$ of the left lead, all transmission matrix elements can be obtained. This sounds actually more complicated than it is. Gaussian elimination (LU decomposition) is the most time consuming algorithmic step in solving the linear equations of Eq. 2.45. This step however is independent of $\mathbf{Q}_{L,+,m}$. The latter only enters in the second algorithmic step, i.e. back substitution, which is much cheaper and is performed for each $\mathbf{Q}_{L,+,m}$. Moreover, the back substitution step does not have to be complete, since only the coefficients $\mathbf{c}_{S+1}$ are needed to calculate the transmission.[22]

Although the mode matching formalism allows for calculating $t'_{n,m}$ matrix elements according to Eq. 2.49 between all states n and m, i.e. propagating states and evanescent states, only propagating states contribute to the transmission. In this sense it is allowed to restrict n and m to the propagating states only. This does not mean that one can throw out the evanescent states altogether, though. One needs to include them in properly matching the scattering region to the leads, cf. Eq. 2.48. In other words, the Bloch matrices, Eq. 2.37, need to be constructed from all N modes, propagating and evanescent.[23] Both of them are needed for a complete basis.

2.2.3   Green function expressions

For those of you who cannot live without Green functions, I can give you the Landauer formalism expressed in Green functions. The derivations are rather technical, so I will skip most of them. You can find them in the literature. With respect to the Hamiltonian matrix of Eq. 2.46 a finite **Green function matrix** can be defined as

$$\mathbf{G}(E) = (E\mathbf{1} - \mathbf{H}')^{-1}. \tag{2.51}$$

It can be calculated by matrix inversion using essentially the same block Gaussian elimination scheme as in solving the set of linear equations, Eq. 2.45. The matrix $\mathbf{H}'$ is non-Hermitian; its eigenvalues are not real, so the Green function matrix can be evaluated for real energies.[24]

With respect to the original infinite Hamiltonian $\mathbf{H}$ of Eq. 2.27 one can define the usual retarded (infinite) Green function matrix

$$\mathbf{G}^r(E) = [(E + i\eta)\mathbf{1} - \mathbf{H}]^{-1}, \tag{2.52}$$

---

[21]Again, replace n by $\mathbf{k}_{\parallel}n$ or $\mathbf{k}_{\parallel}n\sigma$, when appropriate.

[22]However, if one wants to have the full wave function ψ or calculate the reflection coefficients, the back transformation has to be complete. It is not expensive, though (compared to the Gaussian elimination step).

[23]Only the full set forms a complete basis.

[24]The exceptions are truly bound states of $\mathbf{H}'$. These occur at real energies, at which the Green function matrix thus has poles. However, these occur outside the energy range that is of interest for scattering. At bound state energies only evanescent modes exist in the leads, so this situation is easily identified in practice.

where one needs the infinitesimal η to avoid the poles on the real axis. The advanced Green function $\mathbf{G}^a(E)$ can be obtained from Eq. 1.80. One can show that for layers in the scattering region one has

$$\mathbf{G}_{i,j}(E) = \mathbf{G}_{i,j}^r(E); \ i,j = 0,\dots,S+1. \tag{2.53}$$

if one properly takes the $\lim_{\eta\to 0}$. In terms of the Green function, the wave function in the scattering region can be written as

$$\boldsymbol{\psi} = \mathbf{G}(E)\,\mathbf{Q}_{L,+,m}, \tag{2.54}$$

and from Eq. 2.49 one obtains

$$t_{n,m} = \sqrt{\frac{v_{L,+,m}}{v_{R,+,n}}}\ \widetilde{\mathbf{u}}_{R,+,n}^\dagger\ \mathbf{G}_{S+1,0}(E)\,\mathbf{Q}_{L,+,m}. \tag{2.55}$$

After some manipulation one can obtain a **Fisher-Lee expression** that generalizes the 1D expression of Eq. 1.83

$$t_{n,m} = i\hbar\sqrt{\frac{v_{R,+,n}v_{L,+,m}}{a_L a_R}}\ \widetilde{\mathbf{u}}_{R,+,n}^\dagger\ \mathbf{G}_{S+1,0}(E)\,\widetilde{\mathbf{u}}_{L,+,m}'. \tag{2.56}$$

The notation is a bit messy due to the fact the modes $\mathbf{u}_{L/R,+,m/n}$ are not orthogonal; see Sec. 2.2.1. If they were orthogonal, then the expression would simply contain $\mathbf{u}_{R,+,n}^\dagger\mathbf{G}_{S+1,0}(E)\mathbf{u}_{L,+,m}$. This indicates that the transmission amplitude from mode $\mathbf{u}_{L,+,m}$ in the left lead to mode $\mathbf{u}_{R,+,n}$ in the right lead is determined by the Green function matrix that brings you from layer $i = 0$ to layer $i = S+1$, i.e. across the scattering region. Proper orthogonalization of the modes leads to expressions for $\widetilde{\mathbf{u}}_{R,+,n}^\dagger$ and $\widetilde{\mathbf{u}}_{L,+,m}'$. The details are a bit messy and can be found in the literature. The Bloch velocities $v_{L/R,+,m/n}$ serve to make scattering matrix unitary, as before. Since $v_{L/R,+,m/n} \neq 0$ only for propagating states, Eq. 2.56 expresses explicitly that the transmission is zero whenever n or m describes an evanescent mode. The layer thicknesses $a_{L/R}$ are just normalization factors, since our modes are normalized within a layer.

After some algebra the Green function expressions we obtained in the one-dimensional case can be generalized to three dimensions. The velocity of Eq. 1.77 becomes a velocity matrix

$$\mathbf{V}_{L/R} = -\frac{2a_{L/R}}{\hbar}\,\mathrm{Im}\,\Sigma_{L/R}. \tag{2.57}$$

It is a diagonal matrix of dimension N (the total number of modes). The diagonal matrix elements are the mode velocities $v_{L/R,\pm,n}$ of Eq. 2.34. Since $v_{L/R,\pm,n} = 0$ for evanescent states, this means that in general the velocity matrices are singular. The transmission amplitudes of Eq. 2.56 can be assembled in a transmission matrix $\mathbf{t}$, which using the velocity matrices can be expressed as

$$\mathbf{t} = 2i\sqrt{-\,\mathrm{Im}\,\Sigma_R}\ \mathbf{G}_{S+1,0}(E)\ \sqrt{-\,\mathrm{Im}\,\Sigma_L}. \tag{2.58}$$

Note that this generalizes Eq. 1.84. Finally, the total transmission can be expressed as

$$T = \mathrm{Tr}\left[\mathbf{t}^\dagger\mathbf{t}\right] = 4\mathrm{Tr}\left[\mathrm{Im}\,\Sigma_R\,\mathbf{G}_{S+1,0}^r(E)\,\mathrm{Im}\,\Sigma_L\,\mathbf{G}_{0,S+1}^a(E)\right] \tag{2.59}$$

which generalizes Eq. 1.85. This is known as the **Caroli expression** or the **NEGF expression**.[25]

---

[25] NEGF = non-equilibrium Green function. In the linear response regime one can use equilirium Green functions, however, which is what we are doing here.