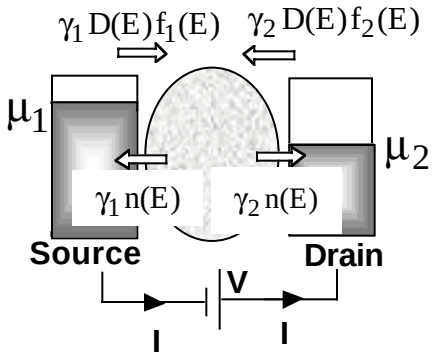
9/ Coherent transport

(The reader may wish to review Section 1.6 before reading this chapter)

9.1. Overview

Since this Chapter is rather long, let me start with a detailed overview, that can also serve as a "summary". In Chapter one, I described a very simple model for current flow, namely a single level ε which communicates with two contacts, labeled the source and the drain. The strength of the coupling to the source (or the drain) was characterized by the rate $\gamma_1/\hbar$ (or $\gamma_2/\hbar$) at which an electron initially occupying the level would escape into the source (or the drain).

Fig.9.1.1. Flux of electrons into and out of a channel: Independent level model, see Eqs.(1.6.4) – (1.6.6)



I pointed out that the flow of current is due to the difference in "agenda" between the source and the drain, each of which is in a state of local equilibrium, but are maintained at two different electrochemical potentials and hence with two distinct Fermi functions

$$f_1(E) \equiv f_0(E-\mu_1) = \frac{1}{\exp\left((E-\mu_1)/k_BT\right)+1} \qquad (9.1.1a)$$

$$f_2(E) \equiv f_0(E-\mu_2) = \frac{1}{\exp\left((E-\mu_2)/k_BT\right)+1} \qquad (9.1.1b)$$

by the applied bias V : $\mu_2 - \mu_1 = -qV$. The source would like the number of electrons occupying the level to be equal to $f_1(\varepsilon)$ while the drain would like to see this number be $f_2(\varepsilon)$. The actual steady state number of electrons, N lies somewhere in-between and the source keeps pumping in electrons while the drain keeps pulling them out, each hoping to establish equilibrium with itself. In the process, a current flows in the external circuit.

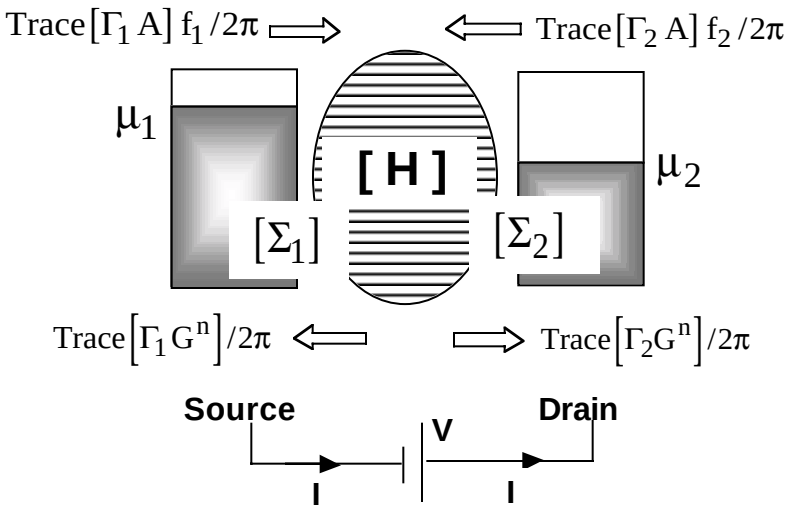


Fig.9.1.2. Inflow and outflow for an arbitrary multi-level device whose energy levels are described by a Hamiltonian matrix [H] and whose coupling to the source and drain contacts is described by self-energy matrices $[\Sigma_1(E)]$ and $[\Sigma_2(E)]$ respectively.

My purpose in this Chapter is essentially to carry out a generalized version of this treatment applicable to an arbitrary multi-level device (Fig.9.1.2) whose energy levels and coupling are described by matrices rather than ordinary numbers:

$\varepsilon \rightarrow [H]$ ***Hamiltonian*** matrix

$\gamma_{1,2} \rightarrow [\Gamma_{1,2}]$ ***Broadening*** matrices, $\Gamma_{1,2} = i\left[\Sigma_{1,2} - \Sigma_{1,2}^{+}\right]$

In Chapter 8 we have seen that connecting a device to a reservoir broadens its energy levels and it is convenient to talk in terms of a continuous independent energy variable E, rather than a discrete set of eigenstates. The density matrix can be written in the form (see Eq.(8.2.13))

$$[\rho] = \int_{-\infty}^{+\infty} (dE/2\pi)\ [G^n(E)] \qquad (91.2a)$$

where in equilibrium $[G^n(E)]_{eq} = [A(E)]\, f_0(E-\mu)$ (9.1.2b)

Just as the spectral function [A] represents the matrix version of the density of states per unit energy, the correlation function $[G^n]$ is the matrix version of the electron density per unit energy.

Non-equilibrium density matrix: In Section 9.1 the first result we will prove is that when the device is connected to two contacts with two distinct Fermi functions $f_1(E)$ and $f_2(E)$, the density matrix is given by Eq.(9.1.1) with (dropping the argument 'E' for clarity)

$$[G^n] = [A_1]\, f_1 + [A_2]\, f_2 \qquad (9.1.3)$$

where $A_1 = G\,\Gamma_1\,G^+$ ***and*** $A_2 = G\,\Gamma_2\,G^+$ (9.1.4)

$$G = [EI - H - \Sigma_1 - \Sigma_2]^{-1} \qquad (9.1.5)$$

suggesting that a fraction $[A_1]$ of the spectral function remains in equilibrium with the source Fermi function f_1 while another fraction $[A_2]$ remains in equilibrium with the drain Fermi function f_2. We will show that these two partial spectral functions indeed add up to give the total spectral function [A] that we discussed in Chapter 8 (see Eq.(8.2.7))

$$[A] \equiv i[G - G^+] = [A_1] + [A_2] \qquad (9.1.6)$$

Current: Next we will show that the current I_i at terminal 'i' can be written in the form

$$I_i = (-q/h) \int_{-\infty}^{+\infty} dE\ \tilde{I}_i(E)$$

with $\tilde{I}_i = \text{Trace}\,[\Gamma_i A]\, f_1 - \text{Trace}\,[\Gamma_i G^n]$ (9.1.7)

representing a dimensionless current per unit energy. This leads to the picture shown in Fig.9.1.2 which can be viewed as the quantum version of our elementary picture from Chapter one (Fig.9.1.1).

One-level model: In Chapter one, we went through an example with just one level so that the electron density and current could all be calculated from a rate equation with a simple model for broadening. I then indicated that in general we need a matrix version of this "scalar model" and that is what the rest of the book is about (see Fig.1.6.5).

It is instructive to check that the full "matrix model" we have stated above (and will derive in this Chapter) reduces to our old results (Eqs.(1.6.4) – (1.6.6)) when we specialize to a one-level system so that all the matrices reduce to pure numbers.

From Eq.(9.1.5), $G(E) = (E - \varepsilon + (i\Gamma/2))^{-1}$

From Eq.(9.1.4), $A_1(E) = \dfrac{\Gamma_1}{(E-\varepsilon)^2 + (\Gamma/2)^2}$, $A_2(E) = \dfrac{\Gamma_2}{(E-\varepsilon)^2 + (\Gamma/2)^2}$

From Eq.(9.1.6) $A(E) = \dfrac{\Gamma}{(E-\varepsilon)^2 + (\Gamma/2)^2}$

From Eq.(9.1.3), $G^n(E) = A(E)\left(\dfrac{\Gamma_1}{\Gamma} f_1(E) + \dfrac{\Gamma_2}{\Gamma} f_2(E)\right)$

which can be compared with Eq.(1.6.4). Similarly, from Eqs.(9.7) the current at the two terminals are given by (cf. Eqs.(1.6.5a,b)):

$$I_1 = \frac{q}{h}\int_{-\infty}^{+\infty} dE\ \Gamma_1 \left[A(E) f_1(E) - G^n(E)\right]$$

$$I_2 = \frac{q}{h}\int_{-\infty}^{+\infty} dE\ \Gamma_2 \left[A(E) f_2(E) - G^n(E)\right]$$

Transmission: Eq.(9.7) can be combined with (9.3) and (9.6) to write

$$\bar{I}_1 = -\bar{I}_2 = \bar{T}(E)\,(f_1(E) - f_2(E))$$

$$\text{where } \bar{T}(E) \equiv \text{Trace}\,[\Gamma_1 A_2] = \text{Trace}\,[\Gamma_2 A_1] \qquad (9.1.8)$$

The current I in the external circuit is given by

$$I = (q/h)\int_{-\infty}^{+\infty} dE\ \bar{T}(E)\ (f_1(E) - f_2(E)) \qquad (9.1.9)$$

The quantity $\overline{T}(E)$ appearing in the current equation (Eq.(9.1.9)) is called the ***transmission*** function which tells us the rate at which electrons transmit from the source to the drain contacts by propagating through the device. Knowing the device Hamiltonian [H] and its coupling to the contacts described by the self-energy matrices $\Sigma_{1,2}$, we can calculate the current either from Eq.(9.1.7) or from Eq.(9.1.9). This procedure can be used to analyze any device as long as the evolution of electrons through the device is ***coherent***. Let me explain what that means.

The propagation of electrons is said to be coherent, if it does not suffer phase-breaking scattering processes that cause a change in the state of an external object. For example, if the electron got deflected from a rigid (that is unchangeable) defect in the lattice, the propagation would still be considered coherent. The effect could be incorporated through an appropriate defect potential in the Hamiltonian [H] and we could still calculate the current from Fig.9.1.2. But, if the electron transferred some energy to the atomic lattice causing it to start vibrating that would constitute a phase-breaking process and the effect cannot be included in [H]. How it can be included is the subject of Chapter 10.

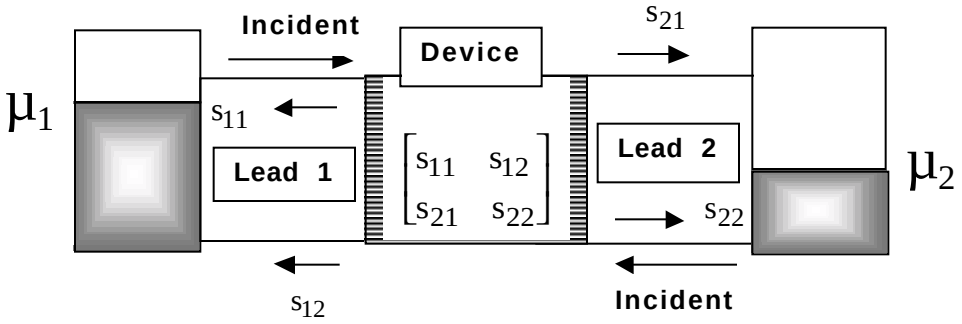


Fig. 9.1.3. The transmission formalism assumes the device to be connected via ideal multimoded quantum wires to the contacts and the transmission function is related to the S-matrix between these leads.

I should mention here that coherent transport is commonly treated using the transmission formalism which starts with the assumption that the device is connected to the contacts by two ideal leads which can be viewed as multimoded quantum wires so that one can calculate an S-matrix for the device (Fig.9.1.3), somewhat like a microwave waveguide. The transmission matrix s_{21} (or s_{12}) is of size M x N (or N x M) if lead 1 has N modes and lead 2 has M modes and the transmission function is obtained from its trace: $\overline{T}(E)$ = Trace $[s_{12} s_{12}^{+}]$ = Trace $[s_{21} s_{21}^{+}]$. This approach is widely used and seems quite appealing especially to those familiar with the concept of S-matrices in microwave waveguides.

Transmission from Green's function: For coherent transport, one can calculate the transmission from the Green's functions method, using the relation

$$\bar{T}(E) \equiv \text{Trace}\left[\Gamma_1 G \Gamma_2 G^+\right] = \text{Trace}\left[\Gamma_2 G \Gamma_1 G^+\right] \qquad (9.1.10)$$

obtained by combining Eq.(9.1.8) with (9.1.4). In Sections 9.2, 9.3 we will derive all the equations (Eq.(9.1.2)-(9.1.7)) stated in this Section. But for the moment let me just try to justify the expression for the transmission (Eq.(9.1.10)) using a simple example. Consider now a simple 1-D wire modeled with a discrete lattice (Fig.9.1.4). We wish to calculate the transmission coefficient

$$\bar{T}(E) = (v_2/v_1)|t|^2 \qquad (9.1.11)$$

where the ratio of velocities (v_2/v_1) is included because the transmission is equal to the ratio of the transmitted to the incident current, and the current is proportional to the velocity times the probability $|\psi|^2$.

To calculate the transmission from the Green's function approach, we start from the Schrodinger equation, $[EI - H]\{\psi\} = \{0\}$, describing the entire infinite system and use the same approach described in Section 8.1 to eliminate the semi-infinite leads

$$[EI - H - \Sigma_1 - \Sigma_2]\{\psi\} = \{S\} \quad \rightarrow \quad \{\psi\} = [G]\{S\} \qquad (9.1.12)$$

where [G] is given by Eq.(9.1.6). Σ_1 and Σ_2 are matrices that represent the effects of the two leads: each has only one non-zero element (see Eq.(8.1.7a)):

$$\Sigma_1(1,1) = -t_0 \exp(ik_1 a), \quad \Sigma_2(N,N) = -t_0 \exp(ik_2 a)$$

corresponding to the end point of the channel (1 or N) where the lead is connected. The source term {S} is a column vector with just one non-zero element corresponding to the end point (1) on which the electron wave is incident (see Eq.(8.1.7b)):

$$S(1) = i2t_0 \sin k_1 a = i(\hbar v_1/a)$$

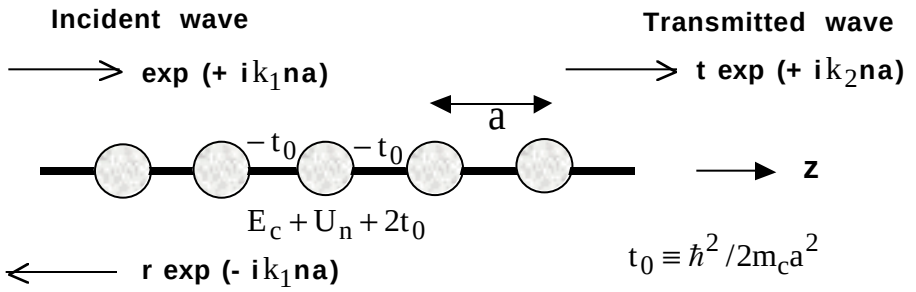


Fig.9.1.4

From Eq.(9.1.12) we can write $t = \psi(N) = G(N,1)\,S(1)$ so that from Eq.(9.1.11)

$$\overline{T}(E) \;=\; (\hbar v_1/a)(\hbar v_2/a)\,\left|G(1,N)\right|^2$$

which is exactly what we get from the general expression in Eq.(9.1.10).

This simple example is designed to illustrate the relation between the Green's function and transmission points of view. I believe the advantages of the Green's function formulation are threefold:

(1) The generality of the derivation shows that the basic results apply to arbitrary shaped channels described by [H] with arbitrary shaped contacts described by $[\Sigma_1]$, $[\Sigma_2]$. This partitioning of the channels from the contacts is very useful when dealing with more complicated structures.

(2) The Green's function approach allows us to calculate the density matrix (hence the electron density) as well. This can be done within the transmission formalism, but less straightforwardly [Ref.9.3].

(3) The Green's function approach can handle incoherent transport with phase-breaking scattering, as we will see in Chapter 10. Phase-breaking processes can only be included phenomenologically within the transmission formalism [Ref.9.2].

In Sections 9.2, 9.3 we will derive the Green's function equations stated earlier (Eqs.(9.1.2)-(9.1.7)), discuss the relation with the transmission formalism in Section 9.4 and finally present a few illustrative examples in Section 9.5.

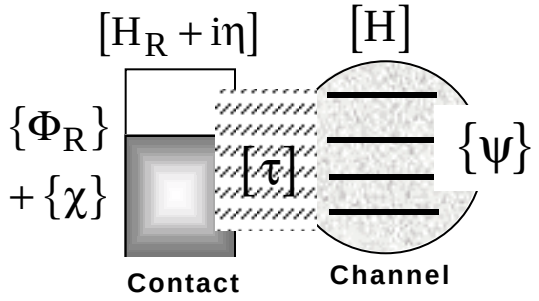
9.2. Density matrix

In this Section we will derive the results stated in Section 9.1 for the non-equilibrium density matrix (Eqs.(9.1.3)-(9.1.6)) for a channel connected to two contacts. In the next Section we will derive the current expressions (Eqs.(9.1.7)-(9.1.9)).

Channel with one contact: I would like to start by re-visiting the problem of a channel connected to one contact and clearing up a conceptual issue, before we take on the real problem with two contacts. In Section 8.1 we started from a Schrodinger equation for the composite contact-channel system

$$E\begin{Bmatrix} \psi \\ \Phi_R + \chi \end{Bmatrix} = \begin{bmatrix} H & \tau \\ \tau^+ & H_R + i\eta \end{bmatrix}\begin{Bmatrix} \psi \\ \Phi_R + \chi \end{Bmatrix} \quad (9.2.1)$$

and showed that the scattered waves $\{\psi\}$ and $\{\chi\}$ can be viewed as arising from the "spilling over" of the wavefunction $\{\Phi_R\}$ in the isolated contact. Using straightforward matrix algebra we obtained



$$\{\chi\} = G_R\,\tau^+\{\psi\} \quad (9.2.2)$$

where $G_R \equiv [EI_R - H_R + i\eta]^{-1}$ (9.2.3)

$$\{\psi\} = [G]\{S\} \quad (9.2.4)$$

$$G \equiv [EI - H - \Sigma]^{-1} \quad (9.2.5)$$

$$\Sigma \equiv \tau\, G_R\, \tau^+ \quad (9.2.6)$$

$$\{S\} = \tau\{\Phi_R\} \quad (9.2.7)$$

Since there is only one contact this is really an equilibrium problem and the density matrix is obtained simply by filling up the spectral function

$$A(E) = i\left[G - G^+\right] \quad (9.2.8)$$

according to the Fermi function as stated in Eqs.(9.1.1), (9.1.2). What I would like to do now is to obtain this result in a completely different way. I will assume that the source waves $\{\Phi_R\}$ from the contact are filled according to the Fermi function and the channel itself is

filled simply by the spilling over of these wavefunctions. I will show that the resulting density matrix in the channel is identical to what we obtained earlier. Once we are clear about the approach we will extend it to the real problem with two contacts.

Before we connect the contact to the device, the electrons will occupy the contact eigenstates α according to its Fermi function, so that we can write down the density matrix for the contact as

$$\rho_R(\vec{r},\vec{r}') = \sum_\alpha \phi_\alpha(\vec{r})\, f_0(\varepsilon_\alpha - \mu)\, \phi_\alpha^*(\vec{r}')$$

or in matrix notation as $$[\rho_R] = \sum_\alpha f_0(\varepsilon_\alpha - \mu)\{\phi_\alpha\}\{\phi_\alpha\}^+ \qquad (9.2.9)$$

Now we wish to calculate the device density matrix by calculating the response of the device to the excitation $\tau\{\Phi\}$ from the contact. We can write the source term due to each contact eigenstate α, $\{S_\alpha\} = \tau\{\phi_\alpha\}$

find the resulting device wavefunction from Eq.(9.2.8) $\{\psi_\alpha\} = G\,\tau\{\phi_\alpha\}$

and then obtain the device density matrix by adding up the individual components weighted by the appropriate Fermi factors for the original contact eigensate α:

$$\begin{aligned}
[\rho] &= \sum_\alpha f_0(\varepsilon_\alpha - \mu)\{\psi_\alpha\}\{\psi_\alpha\}^+ \\
&= \int dE\, f_0(E-\mu) \sum_\alpha \delta(E-\varepsilon_\alpha)\{\psi_\alpha\}\{\psi_\alpha\}^+ \\
&= \int dE\, f_0(E-\mu)\, G\,\tau \left[\sum_\alpha \delta(E-\varepsilon_\alpha)\{\phi_\alpha\}\{\phi_\alpha\}^+\right]\tau^+ G^+ \\
&= \int \frac{dE}{2\pi} f_0(E-\mu)\, G\,\tau\, A_R\, \tau^+ G^+ \qquad (9.2.10)
\end{aligned}$$

making use of the expression for the spectral function in the contact (Eq.(8.2.3):

$$A_R(E) = \sum_\alpha \delta(E-\varepsilon_\alpha)\{\phi_\alpha\}\{\phi_\alpha\}^+ \qquad (9.2.11)$$

From Eq.(9.2.6), $$\Gamma = i[\Sigma - \Sigma^+] = \tau A_R \tau^+ \qquad (9.2.12)$$

so that from Eq.(9.2.10) we can write

$$[G^n] = [G\,\Gamma\,G^+]\,f_0(E-\mu) \qquad (9.2.13)$$

where we have made use of Eq.(9.2.2). To show that this is the same as our earlier result (Eqs.(9.1.1), (9.1.2)), we need the following ***important*** identity:

$$\text{If} \quad [G] = [EI-H-\Sigma]^{-1} \quad \text{and} \quad \Gamma = i\left[\Sigma-\Sigma^+\right]$$
$$\text{Then} \quad A \equiv i\left[G-G^+\right] = G\,\Gamma\,G^+ = G^+\,\Gamma\,G \qquad (9.2.14)$$

This is shown by writing $\left(G^+\right)^{-1} - G^{-1} = \Sigma-\Sigma^+ = -i\,\Gamma$

then pre-multiplying with G and post-multiplying with G^+ to obtain

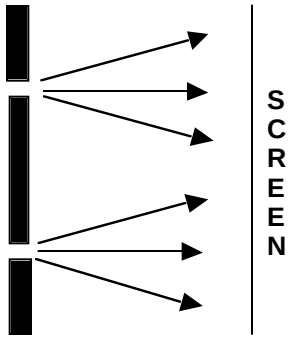
$$G - G^+ = -i\,G\,\Gamma\,G^+ \rightarrow A = G\,\Gamma\,G^+$$

Alternatively if we pre-multiply with and post-multiply with G we obtain

$$G - G^+ = -i\,G^+\,\Gamma\,G \rightarrow A = G^+\,\Gamma\,G$$

I have used this simple one contact problem to illustrate the important physical principle that the different eigenstates are ***uncorrelated*** and so we should calculate their contributions to the density matrix independently and then add them up.

This is a little bit like Young's two-slit experiment. If the two slits are illuminated coherently then the intensity on the screen will show an interference pattern. But if the slits are illuminated incoherently then the intensity on the screen is simply the sum of the intensities we would get from ***each slit independently***. Each eigenstate α is like a "slit" that "illuminates" the device and the important point is that the "slits" have no phase coherence.



That is why we calculate the device density matrix for each 'slit" α independently and add them up.

Now that we have been through this exercise once, it is convenient to devise the following rule for dealing with the contact and channel wavefunctions

$$\{\Phi_R\}\{\Phi_R^+\} \Rightarrow \int (dE/2\pi)\, f_0(E-\mu)\,[A_R(E)] \quad (9.2.15a)$$

$$\{\psi\}\{\psi^+\} \Rightarrow \int (dE/2\pi)\,[G^n(E)] \quad (9.2.15b)$$

reflecting the fact that the electrons in the contact are distributed according to the Fermi function $f_0(E-\mu)$ in a continuous distribution of eigenstates described by the spectral function, $A_R(E)$. This rule can be used to shorten the algebra considerably. For example, to evaluate the density matrix we first write down the result for a single eigenstae

$$\{\psi\} = G\,\tau\{\Phi_R\} \rightarrow \{\psi\}\{\psi\}^+ = G\,\tau\{\Phi_R\}\{\Phi_R\}^+\,\tau^+ G^+$$

and then apply Eq.(9.2.15) to obtain $[G^n] = [G\,\tau A_R\,\tau^+ G^+]\, f_0(E-\mu)$ which reduces to Eq.(9.2.13) making use of Eq.(9.2.12).

Channel with two contacts: Now we are ready to tackle the actual problem with two contacts. We assume that before connecting to the channel, the electrons in the source and the drain contact have wavefunctions $\{\Phi_1\}$, $\{\Phi_2\}$ obeying the Schrodinger equations for the isolated contacts:

$$[E\,I - H_1 + i\eta]\{\Phi_1\} = \{0\} \quad \text{and} \quad [E\,I - H_2 + i\eta]\{\Phi_2\} = \{0\} \quad (9.2.16)$$

where $[H_1], [H_2]$ are the Hamiltonians for contacts 1 and 2 respectively and we have added a small positive infinitesimal times an identity matrix, $[\eta] = 0^+\,[I]$, to introduce dissipation as before. When we couple the device to the contacts as shown in Fig.9.2.1, these electronic states from the contacts "spill over" giving rise to a wavefunction $\{\psi\}$ inside the device which in turn excites scattered waves $\{\chi_1\}$ and $\{\chi_2\}$ in the source and drain respectively.

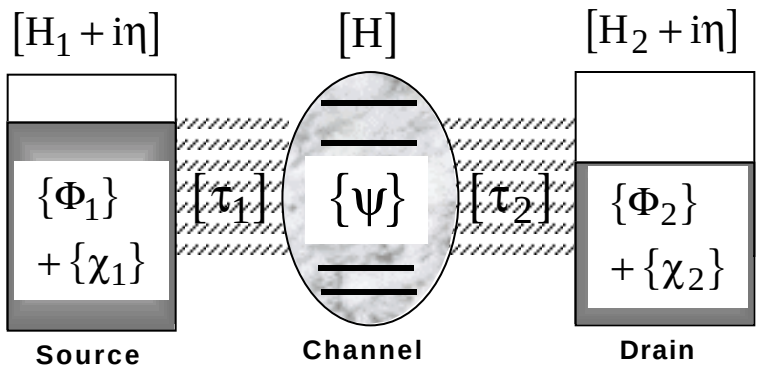


Fig.9.2.1. A channel connected to two contacts.

The overall wavefunction will satisfy the composite Schrodinger equation for the composite contact 1 - device- contact 2 system which we can write in three blocks (cf. Eq.(9.2.1)):

$$\begin{array}{c} \\ \textbf{contact 1} \\ \textbf{device} \\ \textbf{contact 2} \end{array} \begin{array}{c} \begin{array}{ccc} \textbf{contact 1} & \textbf{device} & \textbf{contact 2} \end{array} \\ \begin{pmatrix} EI - H_1 + i\eta & -\tau_1^+ & 0 \\ -\tau_1 & EI - H & -\tau_2 \\ 0 & -\tau_2^+ & EI - H_2 + i\eta \end{pmatrix} \end{array} \begin{Bmatrix} \Phi_1 + \chi_1 \\ \psi \\ \Phi_2 + \chi_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} \quad (9.2.17)$$

where [H] is the channel Hamiltonian. Using straightforward matrix algebra we obtain from the first and last equations

$$\{\chi_1\} = G_1 \tau_1^+ \{\psi\} \quad \textit{and} \quad \{\chi_2\} = G_2 \tau_2^+ \{\psi\} \quad (9.2.18)$$

$$\text{where} \quad G_1 = [EI - H_1 + i\eta]^{-1} \quad \textit{and} \quad G_2 = [EI - H_2 + i\eta]^{-1} \quad (9.2.19)$$

are the Green's functions for the isolated reservoirs. Using Eq.(9.2.18) to eliminate $\{\chi_1\}$,$\{\chi_2\}$ from the middle equation in Eqs.(9.2.2) we obtain

$$[EI - H - \Sigma_1 - \Sigma_2]\{\psi\} = \{S\} \quad (9.2.20)$$

$$\text{where} \quad \Sigma_1 = \tau_1 G_1 \tau_1^+ \quad \text{and} \quad \Sigma_2 = \tau_2 G_2 \tau_2^+ \quad (9.2.21)$$

are the self-energy matrices that we discussed in Chapter 8, The corresponding broadening matrices (Eq.(9.1.1)) are given by

$$\Gamma_1 = \tau_1 A_1 \tau_1^+ \quad \text{and} \quad \Gamma_2 = \tau_2 A_2 \tau_2^+ \quad (9.2.22)$$

where $A_1 = i\left[G_1 - G_1^+\right]$ and $A_2 = i\left[G_2 - G_2^+\right]$ are the spectral functions for the isolated contacts 1 and 2 respectively. Also,

$$\{S\} \equiv \tau_1\{\Phi_1\} + \tau_2\{\Phi_2\} \tag{9.2.23}$$

is the sum of the source terms $\tau_1\Phi_1$ (from the source) and $\tau_2\Phi_2$ (from the drain) as shown in Fig.9.2.2.

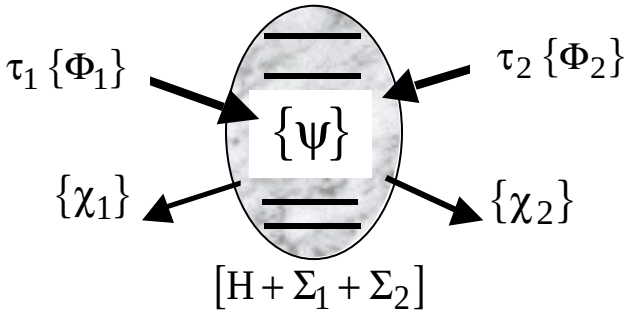


Fig.9.2.2. Channel excited by $\tau_1\Phi_1$ (from the source) and $\tau_2\Phi_2$ (from the drain). The channel response is decribed by Eq.(9.1.20) and it in turn generates $\{\chi_1\}$,$\{\chi_2\}$ in the contacts (see Eq.(9.1.18)).

To evaluate the density matrix, we define the channel Green's function

$$G \equiv [EI - H - \Sigma_1 - \Sigma_2]^{-1} \tag{9.2.24}$$

and use it to express the channel wavefunction in terms of the source terms from Eq.(9.2.20):

$$\{\psi\} = G\{S\} \quad \rightarrow \quad \{\psi\}\{\psi\}^+ = G\{S\}\{S\}^+ G^+ \tag{9.2.25}$$

Note that the cross terms in the source

$$SS^+ = \tau_1\Phi_1\Phi_1^+\tau_1^+ + \tau_2\Phi_2\Phi_2^+\tau_2^+ + \tau_1\Phi_1\Phi_2^+\tau_2^+ + \tau_2\Phi_2\Phi_1^+\tau_1^+$$

are zero since Φ_1 and Φ_2 are the wavefunc **Cross-terms = 0** g to the channel) in the source and drain contacts which are physically disjoint and unconnected. The direct terms

are evaluated using the basic principle (see Eq.(9.2.15)) that we formulated earlier with the one-contact problem:

$$\Phi_1 \Phi_1^+ \Rightarrow \int (dE/2\pi)\, f_1(E)\, A_1(E) \qquad (9.2.26a)$$

$$\Phi_2 \Phi_2^+ \Rightarrow \int (dE/2\pi)\, f_2(E)\, A_2(E) \qquad (9.2.26b)$$

to write down the density matrix from $\{\psi\}\{\psi\}^+ = G\{S\}\{S\}^+ G^+$

$$\rho = \int (dE/2\pi) \left\{ \left[G\tau_1 A_1 \tau_1^+ G^+ \right] f_1 + \left[G\tau_2 A_2 \tau_2^+ G^+ \right] f_2 \right\}$$

Making use of Eq.(9.2.22) we can simplify this expression to write

$$G^n = G\Sigma^{in} G^+ \qquad (9.2.27)$$

$$\text{and} \quad \left[\Sigma^{in}\right] = [\Gamma_1] f_1 + [\Gamma_2] f_2 \qquad (9.2.28)$$

noting that $[\rho] = \int (dE/2\pi)\,[G^n]$ as defined earlier in Eq.(9.1.2). Just as G^n is obtained $\{\psi\}\{\psi\}^+$, Σ^{in} is obtained from $\{S\}\{S\}^+$. One could thus view Eq.(9.2.27) as a relation between the "electron density" in the device created by the source term $\{S\}$ representing the spill-over of electrons from the contacts.

Partial spectral function: Substituting Eq.(9.2.27) into Eq.(9.2.26) we can write

$$\left[G^n\right] = [A_1] f_1 + [A_2] f_2 \qquad (9.2.29)$$

where $A_1 = G\Gamma_1 G^+$ ***and*** $A_2 = G\Gamma_2 G^+$

Comparing this with the equilibrium result (see Eq.(9.1.2)): $[G^n] = [A]\ f_0$, it seems natural to think of the total spectral function [A(E)] as consisting of two parts: $[A_1(E)]$ arising from the spill-over (or propagation) of states in the left contact and $[A_2(E)]$ arising from the spill-over of states in the right contact. The former is filled according to the left Fermi function $f_1(E)$ while the latter is filled according to the right Fermi function $f_2(E)$.

To show that the two partial spectra indeed add up to give the correct total spectral function: $A = A_1 + A_2$, we note from Eq.(9.2.14) that since the self energy Σ has two parts Σ_1 and Σ_2 coming from two contacts, $A = G[\Gamma_1 + \Gamma_2]G^+ = A_1 + A_2$ as stated in Eq.(8.7).

Exclusion principle ? An important conceptual point before we move on. Our approach is to use the Schrodinger equation to calculate the evolution of a specific eigenstate Φ_α from one of the contacts and then superpose the results from distinct eigenstates to obtain the basic rule stated in Eq.(9.2.15) or Eq.(9.2.26). It may appear that by superposing all these individual fluxes we are ignoring the Pauli exclusion principle. Wouldn't the presence of electrons evolving out of one eigenstate ***block*** the flux evolving out of another eigenstate? The answer is no, as long as the evolution of the electrons is coherent. This is easiest to prove in the time domain, by considering two electrons that originate in distinct eigenstates $\{\Phi_1\}$ and $\{\Phi_2\}$. Initially there is no question of one blocking the other since they are orthogonal: $\{\Phi_1\}^+\{\Phi_2\} = 0$. At later times their wavefunctions can be written as

$$\{\psi_1(t)\} = \exp[-iHt/\hbar]\{\Phi_1\}$$
$$\text{and} \quad \{\psi_2(t)\} = \exp[-iHt/\hbar]\{\Phi_2\}$$

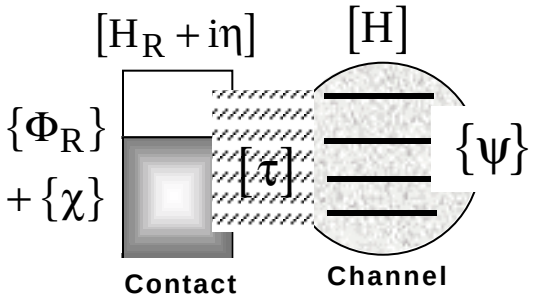
if both states evolve coherently according to the Schrodinger equation: $i\hbar\, d\{\psi\}/dt = [H]\{\psi\}$ It is straightforward to show that the overlap between any two states does not change as a result of this evolution: $\{\psi_1(t)\}^+\{\psi_2(t)\} = \{\Phi_1\}^+\{\Phi_2\}$. Hence wavefunctions originating from orthogonal states remain orthogonal at all times and never "Pauli block" each other. Note, however, that this argument cannot be used when phase-breaking processes (briefly explained in the introduction to this Chapter) are involved since the evolution of electrons cannot be described by a one-particle Schrodinger equation.

9.3. Inflow / Outflow

Now that we have derived the results for the non-equilibrium density matrix (see Eqs.(9.1.3)-(9.1.6)), let us discuss the current flow at the terminals (Eqs.(9.1.7)-(9.1.9)). As before let us start with the "one-contact" problem.

Channel with one contact: Consider again the problem of a channel connected to one contact described by

$$E\begin{Bmatrix}\psi\\ \Phi\end{Bmatrix} = \begin{bmatrix}H & \tau\\ \tau^{+} & H_R + i\eta\end{bmatrix}\begin{Bmatrix}\psi\\ \Phi\end{Bmatrix}$$



which is the same as Eq.(9.2.1) with $\{\Phi\} \equiv \{\Phi_R + \chi\}$ for clarity. How can we evaluate the current flowing between the channel and the contact?

Just as we did in Section 5.4 (when discussing the velocity of a band electron), we need to look at the time-dependent version of this equation

$$i\hbar\frac{d}{dt}\begin{Bmatrix}\psi\\ \Phi\end{Bmatrix} = \begin{bmatrix}H & \tau\\ \tau^{+} & H_R + i\eta\end{bmatrix}\begin{Bmatrix}\psi\\ \Phi\end{Bmatrix}$$

and obtain an expression for the time rate of change in the probability density inside the channel which is given by Trace[$\psi\psi^{+}$] = Trace[$\psi^{+}\psi$] = $\psi^{+}\psi$ (note that $\psi^{+}\psi$ is just a number and so it does not matter if we take the trace or not):

$$I \equiv \frac{d}{dt}\psi^{+}\psi = \frac{\text{Trace}\left[\psi^{+}\tau\Phi - \Phi^{+}\tau^{+}\psi\right]}{i\hbar} \qquad (9.3.1)$$

Noting that $\{\Phi\} \equiv \{\Phi_R + \chi\}$, we can divide this net current, I conceptually into an inflow, proportional to the "incident" wave $\{\Phi_R\}$, and an outflow proportional to the "scattered" wave $\{\chi\}$:

$$I = \underbrace{\frac{\text{Trace}\left[\psi^{+}\tau\Phi_{R} - \Phi_{R}^{+}\tau^{+}\psi\right]}{i\hbar}}_{\text{Inflow}} - \underbrace{\frac{\text{Trace}\left[\chi^{+}\tau^{+}\psi - \psi^{+}\tau\chi\right]}{i\hbar}}_{\text{Outflow}} \qquad (9.3.2)$$

Making use of Eq.(9.2.4) and (9.2.7) we can write the inflow as

$$\text{Inflow} = \text{Trace}\left[S^{+}G^{+}S - S^{+}GS\right]/i\hbar = \text{Trace}\left[SS^{+}A\right]/\hbar$$

since $i\,[G - G^{+}] \equiv [A]$. To obtain the total inflow we need to sum the inflow due to each contact eigenstate α all of which as we have seen is taken care of by the replacement (see Eq.(9.2.15))

$$\Phi_{R}\,\Phi_{R}^{+} \Rightarrow \int \frac{dE}{2\pi} f_{0}(E-\mu)\, A_{R}(E)$$

Since $S = \tau\,\Phi_{R}$, this leads to

$$SS^{+} \Rightarrow \int \frac{dE}{2\pi} f_{0}(E-\mu)\, \tau A_{R}\tau^{+} = \int \frac{dE}{2\pi} f_{0}(E-\mu)\,[\Gamma]$$

so that the inflow term becomes

$$\text{Inflow} = \frac{1}{\hbar}\int \frac{dE}{2\pi} f_{0}(E-\mu)\,\text{Trace}[\Gamma A] \qquad (9.3.3a)$$

Similarly we make use of Eqs.(9.2.2) and (9.2.12) to write the outflow term as

$$\text{Outflow} = \text{Trace}\left[\psi^{+}\tau G_{R}^{+}\tau^{+}\psi - \psi^{+}\tau G_{R}\tau^{+}\psi\right]/i\hbar = \text{Trace}\left[\psi\psi^{+}\Gamma\right]/\hbar$$

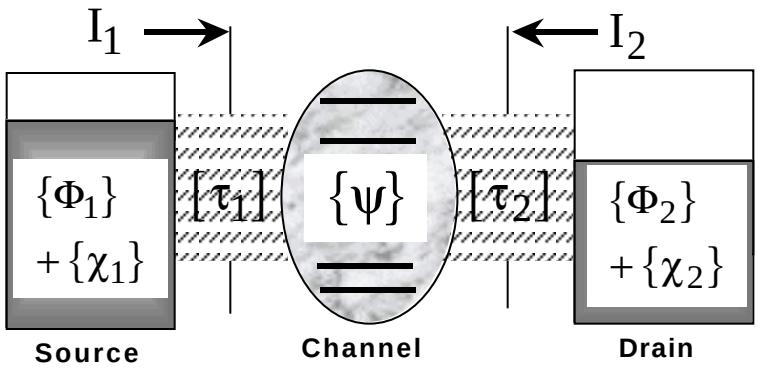
On summing over all the eigenstates $\psi\psi^{+} \Rightarrow \int dE\, G^{n}/2\pi$, so that

$$\text{Outflow} = \frac{1}{\hbar}\int \frac{dE}{2\pi}\,\text{Trace}\left[\Gamma G^{n}\right] \qquad (9.3.3b)$$

It is easy to see that the inflow and outflow are equal at equilibrium, since $G^n = Af_0$ (see Eq.(9.1.2)).

Channel with two contacts: Now we are ready to calculate the inflow and outflow for the channel with two contacts. We consider one of the interfaces, say the one with the source contact, and write the inflow as (cf. Eq.(9.3.2))

$$I_1 = \underbrace{\frac{\text{Trace}\left[\psi^+\tau_1\Phi_1 - \Phi_1^+\tau_1^+\psi\right]}{i\hbar}}_{\text{Inflow}} - \underbrace{\frac{\text{Trace}\left[\chi_1^+\tau_1^+\psi - \psi^+\tau_1\chi_1\right]}{i\hbar}}_{\text{Outflow}}$$



Making use of the relations $\psi = G\,S$ (S, G defined in Eqs.(9.2.23) and (9.2.24)) and $\{S_1\} \equiv \tau_1\{\Phi_1\}$, we can write

$$\text{Inflow} = \text{Trace}\left[S^+G^+S_1 - S_1^+GS\right]/i\hbar = \text{Trace}\left[S_1S_1^+A\right]/\hbar$$

since $S = S_1 + S_2$ and $S_1^+S_2 = S_2^+S_1 = 0$.

Next we sum the inflow due to each contact eigenstate α all of which is taken care of by the replacement (see Eq.(9.2.26))

$$\{\Phi_1\}\{\Phi_1^+\} \quad \Rightarrow \quad \int (dE/2\pi)\, f_1(E)\, [A_1(E)]$$

leading to $\{S_1\}\{S_1^+\} = [\tau_1 \Phi_1 \Phi_1^+ \tau_1^+]$

$$\Rightarrow \quad \int (dE/2\pi)\, [\tau_1 A_1 \tau_1^+]\, f_1(E) \;=\; \int (dE/2\pi)\, [\Gamma_1]\, f_1(E)$$

so that
$$\text{Inflow} \;=\; \frac{1}{\hbar} \int (dE/2\pi)\, f_1(E)\, \text{Trace}\left[\Gamma_1 A\right] \qquad (9.3.4a)$$

Similarly we make use of Eqs.(9.2.18) and (9.2.22) to write the outflow term as

$$\text{Outflow} \;=\; \text{Trace}\left[\psi^+ \tau_1 G_1^+ \tau_1^+ \psi - \psi^+ \tau_1 G_1 \tau_1^+ \psi\right]/i\hbar \;=\; \text{Trace}\left[\psi\psi^+ \Gamma_1\right]/\hbar$$

On summing over all the eigenstates $\{\psi\}\{\psi^+\} \;\Rightarrow\; \int (dE/2\pi)\, [G^n]$, so that

$$\text{Outflow} \;=\; (1/\hbar) \int (dE/2\pi)\, \text{Trace}\left[\Gamma_1 G^n\right] \qquad (9.3.4b)$$

The net current I_i at terminal 'i' is given by the difference between the inflow and the outflow (multiplied by the charge ' – q' of an electron) as stated in Eq.(9.1.7)

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

$$\text{with} \quad \tilde{I}_i \;=\; \text{Trace}\left[\Gamma_i A\right] f_1 \;-\; \text{Trace}\left[\Gamma_i G^n\right] \qquad (9.3.5)$$

and illustrated in Fig.9.1.2.

9.4. Transmission

In the last Section we have obtained expressions for the current at each of the contacts which can be expressed as the difference between an inflow and an outflow. In this Section we will express the current in a slightly different form that gives a different perspective to the problem of current flow and helps establish a connection with the transmission formalism widely used in the literature. We start by combining Eq.(9.1.7) with Eqs. (9.1.3) and (9.1.6) to write

$$\bar{I}_1 = \overline{T}_{12}(E)\,[f_1(E)-f_2(E)] \qquad \text{where} \quad \overline{T}_{12}(E) \equiv \text{Trace}\,[\Gamma_1 A_2]$$

$$\text{and} \quad \bar{I}_2 = \overline{T}_{21}(E)\,[f_2(E)-f_1(E)] \qquad \text{where} \quad \overline{T}_{21}(E) \equiv \text{Trace}\,[\Gamma_2 A_1]$$

We expect the currents at the two terminals to be equal and opposite and this is ensured if Trace $[\Gamma_1 A_2]$ = Trace $[\Gamma_2 A_1]$. To show that they are indeed equal, we make use of Eq.(9.2.14) to show that

$$\text{Trace}\,[\Gamma_1 A] = \text{Trace}\,[\Gamma_1 G\Gamma G^+] = \text{Trace}\,[\Gamma G^+\Gamma_1 G] = \text{Trace}\,[\Gamma A_1]$$

Subtracting Trace$[\Gamma_1 A_1]$ from both sides we obtain the desired result that Trace $[\Gamma_1 A_2]$ = Trace $[\Gamma_2 A_1]$, noting that $\Gamma = \Gamma_1 + \Gamma_2$ and A= $A_1 + A_2$). This allows us to write the current as (noting that $2\pi\hbar$ = h)

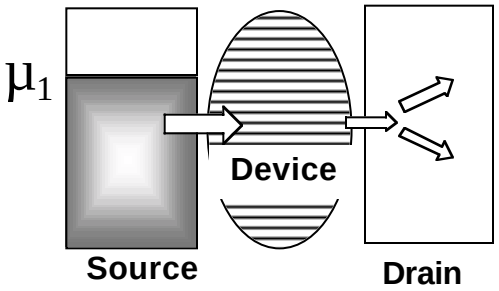
$$I = (q/h)\int_{-\infty}^{+\infty} dE\;\overline{T}(E)\;[f_1(E)-f_2(E)] \tag{9.4.1}$$

where $\overline{T}(E) \equiv$ Trace $[\Gamma_1 A_2]$ = Trace $[\Gamma_2 A_1]$

$$= \text{Trace}\left[\Gamma_1 G \Gamma_2 G^+\right] = \text{Trace}\left[\Gamma_2 G \Gamma_1 G^+\right] \tag{9.4.2}$$

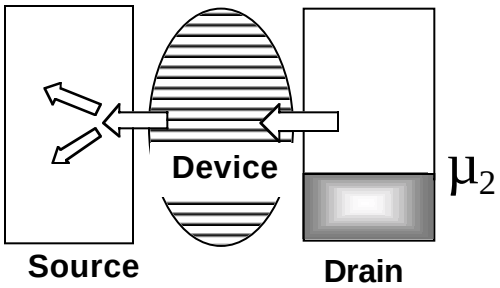
is called the transmission function. Physically we can view the current in Eq.(9.4.1) as the difference between two counterpropagating fluxes, one from the source to the drain and the other from the drain to the source as sketched in Fig.9.4.1. One could view the device as a "semipermeable membrane" that separates two reservoirs of electrons (source and drain) and the transmission function $\overline{T}(E)$ as a measure of the permeability of this membrane to electrons with energy E. We will show that the same function $\overline{T}(E)$ will govern both fluxes at least as long as transport is coherent.

$\bar{I}(S \rightarrow D) = \bar{T}(E) f_1(E)$



(a) Flux of electrons from the source to the drain ($f_2 = 0$)

(b) Flux of electrons from the drain to the source ($f_1 = 0$)



$\bar{I}(D \rightarrow S) = \bar{T}(E) f_2(E)$

Fig.9.4.1. The net current through the device (Eq.(9.2)) can be viewed as the difference between two counterpropagating fluxes from electrons, (a) one from the source to the drain and (b) the other from the drain to the source.

Transmission formalism: In the transmission formalism (sometimes referred to as the Landauer approach) the device is assumed to be connected to the contacts by two uniform leads which can be viewed as quantum wires with multiple modes or subbands (see Chapter 5) having well-defined E-k relationships as sketched in Fig.9.4.1. This allows us to define an S-matrix for the device analogous to a microwave waveguide where the element t_{nm} of the t-matrix tells us the amplitude for an electron incident in mode 'm' in lead 1 to transmit to a mode 'n' in lead 2. It can then be shown that the current is given by Eq.(9.1.1) with the transmission function given by

$$\bar{T}(E) = \sum_m \sum_n |t_{nm}|^2 = \text{Trace}\left[t t^+\right] \qquad (9.4.3)$$

This viewpoint, which is very popular, has the advantage of being based on relatively elementary concepts and also allows one to calculate the transmission function by solving a scattering problem. In the next section we will show with simple examples that this approach yields the same result as that obtained from $\bar{T} = \text{Trace}\,[\Gamma_2 G \Gamma_1 G^+]$ applied to devices with uniform leads

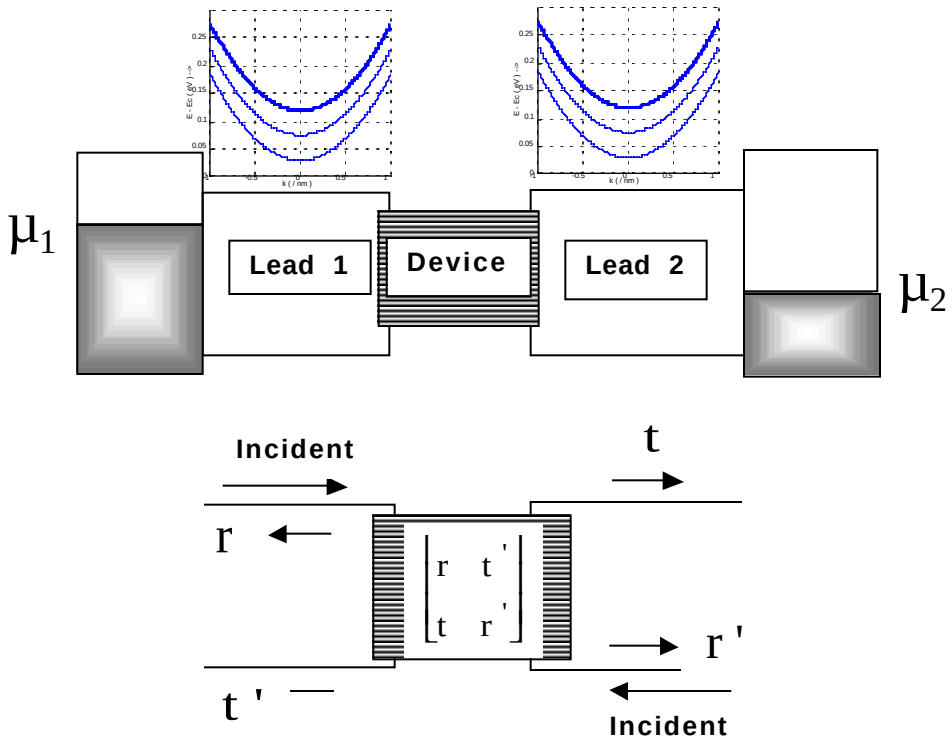


Fig.9.4.1. In the transmission formalism, the device is assumed to be connected to the contacts by two uniform leads which can be viewed as quantum wires with multiple subbands (see Chapter 5) having well-defined E-k relationships as sketched above. This allows us to define an S-matrix for the device analogous to a microwave waveguide.

.

Landauer formula: Landauer pioneered the use of the scattering theory of transport as a conceptual framework for clarifying the meaning of electrical conductance and stressed its fundamental connection to the transmission function: "Conductance ***is*** transmission". This basic relation can be seen starting from Eq.(9.4.1) (making use of Eq.(9.1.1))

$$I = (q/h) \int_{-\infty}^{+\infty} dE\, \bar{T}(E)\, [f_0(E-\mu_1) - f_0(E-\mu_2)]$$

and noting that the current is zero at equilibrium since $\mu_1 = \mu_2$. A small bias voltage V changes each of the functions $\overline{T}$, μ_1 and μ_2 and the resulting current can be written to first order as (δ denotes a small change)

$$I \approx (q/h) \int_{-\infty}^{+\infty} dE\ \delta\overline{T}(E)\ [f_0(E-\mu_1) - f_0(E-\mu_2)]$$

$$+ \quad (q/h) \int_{-\infty}^{+\infty} dE\ \overline{T}(E)\ \delta[f_0(E-\mu_1) - f_0(E-\mu_2)]$$

The first term is zero and the second can be written as

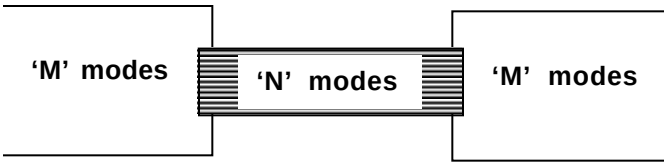
$$I \approx (q^2 V/h) \int_{-\infty}^{+\infty} dE\ \overline{T}(E)\ (-\partial f_0(E)/\partial E)_{E=\mu}$$

so that the conductance is given by

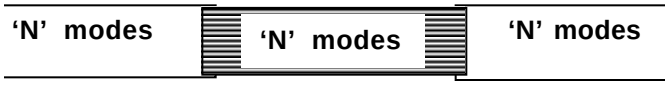
$$G = (q^2/h)\, T_0 \qquad \text{where} \quad T_0 \equiv \int_{-\infty}^{+\infty} dE\ \overline{T}(E)\, F_T(E-\mu) \qquad (9.4.4)$$

and F_T is the thermal broadening function discussed in Chapter 8, which is peaked sharply around $E = \mu$ with a width proportional to $k_B T$ (see Fig.7.3.4). The conductance is thus proportional to the transmission function averaged over an energy range of a few $k_B T$ around the equilibrium electrochemical potential μ, just as the quantum capacitance is proportional to the averaged density of states (cf. Eq.(7.3.8)).

The maximum value of the transmission function (and hence the conductance) is obtained if each of the M subbands or modes in one lead transmits perfectly to the other lead (see Fig.9.4.2). The matrix $[tt^+]$ then is a diagonal matrix of size (MxM) with 1's along the diagonal, so that the transmission is equal to M. This suggests that the maximum transmission is equal to the number of modes 'M' in the leads. But what happens if the device is narrower than the lead and has only 'N' modes, N < M ?



It can be argued that such a structure could not have a transmission any greater than a structure with the leads the same size as the device



since in either case the electrons have to transmit through the narrow device region (assuming that the device is not so short as to allow direct tunneling). Since this latter structure has a maximum transmission of 'N' that must be true of the first structure as well and detailed calculations do indeed show this to be the case. In general we can expect that the maximum transmission is equal to the number of modes in the narrowest segment. Earlier in Chapter 6, we had argued that the maximum conductance of a wire with 'N' modes is equal to $(q^2/h)\,N$ based on the maximum current it could possibly carry.

Buttiker equations: Conductance measurenments are often performed using a four probe structure (Fig.9.4.2) and their interpretation in small structures was initially unclear, till Buttiker came up with an elegant idea [9.1]. He suggested that the Landauer formula

$$G = (q^2/h)\,\tilde{T} \rightarrow I = (q/h)\,\tilde{T}\,[\mu_1 - \mu_2]$$

be extended to structures with multiple terminals by writing the current I_i at the ith terminal as

$$I_i = (q/h)\sum_j \tilde{T}_{ij}\,[\mu_i - \mu_j] \qquad (9.4.5)$$

where $\tilde{T}_{ij}$ is the average transmission from terminal 'j' to 'i'. We know the electrochemical potentials μ at the current terminals (1 and 2), but we do not know them at the voltage terminals which float to a suitable potential so as to make the current zero. How do we calculate the currents from Eq.(9.4.5) since we do not know all the potentials? The point is that of the eight variables (four potentials and four currents), if we know any four, we can calculate the other four with simple matrix algebra. Actually, there are six independent variables. We can always set one of the potentials to zero, since only potential differences give rise to currents. Also, Kirchhoff's law requires all the currents to add up to zero, so that knowing any three currents we can figure out the fourth. So it is convenient to set the potential at one terminal (say #2) equal to zero and write Eq.(9.4.5) in the form of a (3x3) matrix equation.

$$\begin{Bmatrix} I_1 \\ I_3 \\ I_4 \end{Bmatrix} = \frac{q}{h} \begin{bmatrix} \tilde{T}_{12}+\tilde{T}_{13}+\tilde{T}_{14} & -\tilde{T}_{13} & -\tilde{T}_{14} \\ -\tilde{T}_{31} & \tilde{T}_{31}+\tilde{T}_{32}+\tilde{T}_{34} & -\tilde{T}_{34} \\ -\tilde{T}_{41} & -\tilde{T}_{43} & \tilde{T}_{41}+\tilde{T}_{42}+\tilde{T}_{43} \end{bmatrix} \begin{Bmatrix} \mu_1 \\ \mu_3 \\ \mu_4 \end{Bmatrix}$$

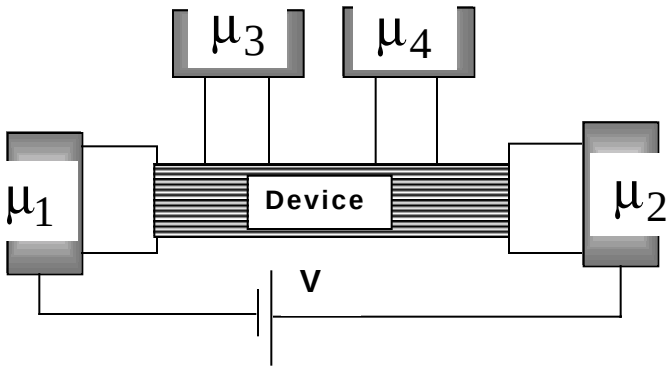


Fig.9.4.2. Conductance measurements are commonly carried out in a four-probe configuration that can be analyzed using the Buttiker equations.

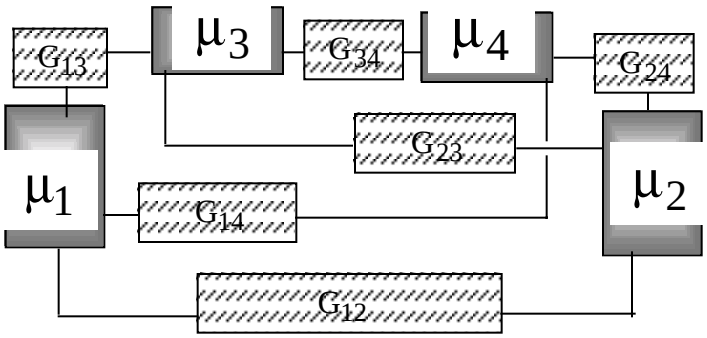
Knowing $\mu_1, I_3 = 0, I_4 = 0$, we can calculate I_1, μ_3, μ_4 and hence the four-probe conductance

$$G_{four-probe} = (\mu_3 - \mu_4)/qI_1$$

We can visualize the Buttiker equations with a simple circuit model if the transmission coefficients are reciprocal, that is, if $\tilde{T}_{ij} = \tilde{T}_{ji}$. These equations are then identical to

Kirchhoff's law applied to a network of conductors $G_{ij} \propto \tilde{T}_{ij} = \tilde{T}_{ji}$ connecting each pair of contacts 'i' and 'j' (see Fig.9.4.3).

Fig.9.4.3. The Buttiker equations can be vcisualized in terms of a conductor network, if the transmission between terminals is reciprocal.



But this picture cannot be used if the transmission coefficients are non-reciprocal: $\tilde{T}_{ij} \neq \tilde{T}_{ji}$, as they are in Hall effect measurements where a magnetic field is present and some of the most notable applications of the Buttiker equations, Eq.(9.4.4), are to the interpretation of such measurements.

Buttiker probes: We have mentioned earlier that the scattering theory of transport can only be used if the electrons transmit coherently through the device so that an S-matrix can be defined. But floating probes effectively extract electrons from the device and reinject them after phase randomization, thus effectively acting as phase-breaking scatterers. This is a seminal observation due to Buttiker which provides a simple phenomenological technique for including the effects of phase-breaking processes in the calculation of current. We simply connect one or more purely conceptual floating probes to the device and then calculate the net current using the Buttiker equations, which can be applied to any number of terminals.

We could even use the general current equation (see Eq.(9.1.1)), rather than the low bias conductance relation) extended to include multiple floating probes:

$$I_i = (q/h) \int_{-\infty}^{+\infty} dE\, \bar{I}_i(E) \qquad (9.4.6)$$

$$\text{where} \quad \bar{I}_i(E) = \sum_j \bar{T}_{ij}(E) \left[f_i(E) - f_j(E) \right] \qquad (9.4.7)$$

One could then adjust the potential μ_j to make the current at each energy equal to zero: $\bar{I}_j(E) = 0$. In principle this could result in different values for μ_j at different energies. Alternatively, we could require a single value for μ_j at all energies which is adjusted to make the total current at all energies equal to zero $\int dE\, \bar{I}_j(E) = 0$. One could then have positive values of $\bar{I}_j(E)$ at certain energies balanced by negative values at other energies making the total come out zero, indicating a flow of electrons from one energy to another due to the scattering processes that the "probe" is expected to simulate. This makes the detailed implementation more complicated since different energy channels get coupled together.

The transmission coefficients at a given energy are usually calculated from the S-matrix for the composite device including the conceptual probes:

$$\bar{T}_{ij} = \text{Trace}\,[s_{ij}(E)\, s_{ij}^{+}(E)]. \qquad (9.4.8)$$

But we could just as well combine this phenomenological approach with our Green's function method using separate self-energy matrices $[\Sigma_i]$ to represent different floating probes and then use the expression

$$\bar{T}_{ij}(E) = \text{Trace}\,[\Gamma_i G \Gamma_j G^{+}] \qquad (9.4.9)$$

to evaluate the transmission. This expression can be derived using the same procedure described earlier for two-terminal structures. The current at terminal 'i' is given by the difference between the inflow and outflow:

$$I_i(E) = (1/\hbar)\ \text{Trace}\ ([\Gamma_i(E)]\ ([A(E)]\, f_i - [G^n(E)]))$$

Making use of the relations $A = \sum_j G\Gamma_j G^{+}$ and $G^n = \sum_j G\Gamma_j G^{+} f_j$

(cf. Eq. (9.2.14)) ***(cf. Eq.(9.2.27))***

we can write $I_i(E) = (1/\hbar) \sum_q \text{Trace}\,[\Gamma_i G \Gamma_j G^+]\,(f_i - f_j)$

so that the current can be written as

$$I_i(E) = (1/\hbar) \sum_j \bar{T}_{ij}\,(f_i - f_j) \qquad (9.4.10)$$

in terms of the transmission function defined above in Eq.(9.4.9).

Sum rule: A very useful result in the scattering theory of transport is the requirement that the sum of the rows or columns of the transmission matrix equals the number of modes:

$$\sum_j \bar{T}_{ij} = \sum_j \bar{T}_{ji} = M_i \qquad (9.4.11)$$

where M_i is the number of modes in lead 'i'. One important consequence of this sum rule is that for a ***two-terminal*** structure $\bar{T}_{12} = \bar{T}_{21}$, even in a magnetic field, since with a (2x2) $\bar{T}$ matrix

$$\begin{bmatrix} \bar{T}_{11} & \bar{T}_{12} \\ \bar{T}_{21} & \bar{T}_{22} \end{bmatrix} \quad \text{we have} \qquad \bar{T}_{11} + \bar{T}_{12} = M_1 = \bar{T}_{11} + \bar{T}_{21} \;\rightarrow\; \bar{T}_{12} = \bar{T}_{21}$$

Note that a similar argument would not work with more than two terminals. For example, with a three terminal structure we could show that $\bar{T}_{12} + \bar{T}_{13} = \bar{T}_{21} + \bar{T}_{31}$, but we could not prove that $\bar{T}_{12} = \bar{T}_{21}$ or that $\bar{T}_{13} = \bar{T}_{31}$.

The Green's function-based expressions for the transmission (see Eq.(9.3.7)) also yield a similar sum rule:

$$\sum_j \bar{T}_{ij} = \sum_j \bar{T}_{ji} = \text{Trace}\,[\Gamma_i A] \qquad (9.4.12)$$

This is shown by noting that

$$\sum_j \overline{T}_{ij} = \sum_j \text{Trace}\left[\Gamma_i G \Gamma_j G^+\right] = \text{Trace}\left[\Gamma_i G \Gamma G^+\right] = \text{Trace}\left[\Gamma_i A\right]$$

where we have made use of Eq.(9.2.14) in the last step. Similarly,

$$\sum_j \overline{T}_{ji} = \sum_j \text{Trace}\left[\Gamma G \Gamma_i G^+\right] = \text{Trace}\left[\Gamma_i G^+ \Gamma G\right] = \text{Trace}\left[\Gamma_i A\right]$$

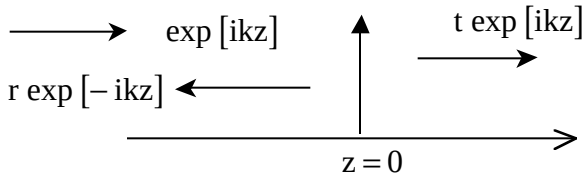
The quantity Trace $[\Gamma_i A]$ thus plays the same role that the number of modes M_i plays in the scattering theory of transport. Interestingly, while M_i is an integer, Trace $[\Gamma_i A]$ can take on any non-integer value. For example, if the device were a really small one having just one level with $E = \varepsilon$, communicating with multiple reservoirs then

$$\Gamma_i A = \frac{\Gamma_i}{(E-\varepsilon)^2 + (\Gamma/2)^2} \qquad \text{with} \qquad \Gamma = \sum_i \Gamma_i$$

which has the shape of a Lorentzian if the broadening is energy-independent. Clearly this can have any fractional value.

9.5. Examples

9.5.1. An analytical example: To see that the Green's function formalism gives the same answer as the scattering theory of transport it is instructive to go through a simple example where the results are easily worked out on paper. Consider for example, a linear conductor with a repulsive potential $U(z) = U_0\,\delta(z)$ at z = 0. The coefficients 'r' and 't' are obtained by requiring that the wavefunction be continuous at z = 0 :



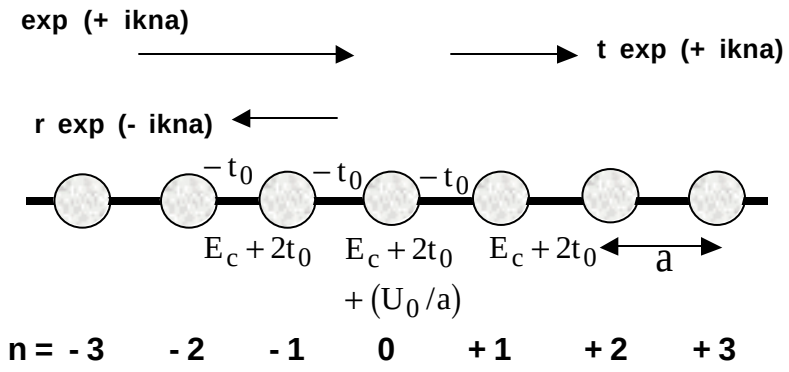
$$[\psi]_{z=0^+} - [\psi]_{z=0^-} = 0 \rightarrow t-(1+r) = 0 \qquad (9.5.1a)$$

and that the derivative be discontinuous by

$$\left[\frac{d\psi}{dz}\right]_{z=0^+} - \left[\frac{d\psi}{dz}\right]_{z=0^-} = \frac{2mU_0}{\hbar^2}[\psi]_{z=0} \rightarrow ik\left[t-(1-r)\right] = \frac{2mU_0 t}{\hbar^2} \qquad (9.5.1b)$$

Eqs.(9.5.1a,b) are solved to yield $t = \dfrac{i\hbar v}{i\hbar v - U_0} \rightarrow \overline{T} = |t|^2 = \dfrac{\hbar^2 v^2}{\hbar^2 v^2 + U_0^{\,2}}$

Finite difference method: Let us now re-do this problem using a discrete lattice with points spaced by 'a', the central cell having an extra potential (U_0/a) for the delta function.



$$t_0 \equiv \hbar^2/2m_c a^2$$

We can carry out a discrete lattice version of the calculation described above, starting from

$$E\psi_0 = \left(E_c + 2t_0 + (U_0/a)\right)\psi_0 \;-\; t_0\psi_{-1} \;-\; t_0\psi_{+1} \tag{9.5.2}$$

and then writing $\psi_0 = 1 + r = t$

$$\psi_{+1} = t\exp[+ika]$$

$$\psi_{-1} = \exp[-ika] + r\exp[+ika]$$

so that $\psi_{+1} = \psi_0\exp[+ika]$

$$\psi_{-1} = -2i\sin[ka] + \psi_0\exp[+ika] \tag{9.5.3}$$

Substituting back into Eq.(9.5.2), we have

$$\left(E - E_c - 2t_0 - [U_0/a] + 2t_0\exp[+ika]\right)\psi_0 = 2i\,t_0\sin ka$$

Making use of the dispersion relation

$$E = E_c + 2t_0(1-\cos ka) \;\rightarrow\; \hbar v(E) = 2at_0\sin ka \tag{9.5.4}$$

this is simplified to $\left(-[U_0/a] + 2i\,t_0\sin[ka]\right)\psi_0 = 2i\,t_0\sin ka$

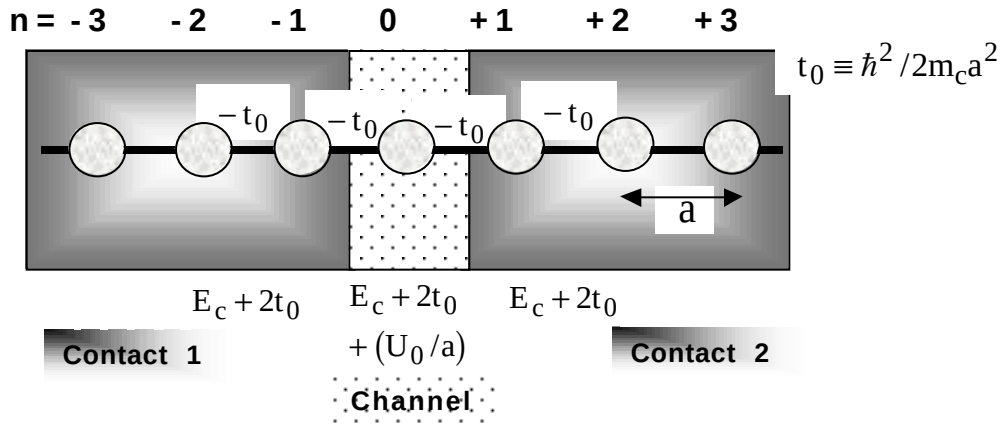
that is, $$\psi_0 = \frac{i\hbar v}{i\hbar v - U_0} \tag{9.5.5}$$

Hence the transmission is given by

$$\overline{T}(E) = |t|^2 = |\psi_0|^2 = \frac{\hbar^2 v(E)^2}{\hbar^2 v(E)^2 + U_0^2} \tag{9.5.6}$$

Green's function method: Finally let us do this problem using the Green's function formulation presented in this Chapter. We treat just one point as the 'device' with a (1x1) Hamiltonian given by

$$[H] = E_c + 2t_0 + (U_0/a)$$



while the effects of the two semi-infinite leads (one on each side) are represented by (1x1) self-energy matrices as discussed in Chapter 8:

$$[\Sigma_1(E)] = -t_0 \exp(ika) \qquad \textbf{\textit{and}} \qquad [\Sigma_2(E)] = -t_0 \exp(ika)$$

where 'ka' is related to the energy 'E' by the dispersion relation (see Eq.(9.5.4)), so that

$$[\Gamma_{1,2}(E)] = i\left[\Sigma_{1,2} - \Sigma_{1,2}^{+}\right] = 2t_0 \sin ka = \hbar v/a$$

Since all matrices are (1x1) in size, it is easy to write down the Green's function:

$$G = [EI - H - \Sigma_1 - \Sigma_2]^{-1}$$
$$= [E - E_c - 2t_0 + 2t_0 \exp(ika) - (U_0/a)]^{-1}$$

Using the dispersion relation to simplify as before

$$G = [i2t_0 \sin ka - (U_0/a)]^{-1} = a/(i\hbar v - U_0)$$

so that the transmission is given by

$$\overline{T}(E) = \text{Trace}\left[\Gamma_1 G \Gamma_2 G^{+}\right] = \frac{\hbar^2 v(E)^2}{\hbar^2 v(E)^2 + U_0^2}$$

in agreement with the earlier result.

9.5.2. Numerical example: The real power of the Green's function method, of course lies not in simple problems like this, but in its ability to handle complex problems without the need for any additional formulation or setting up. Given a Hamiltonian [H] and self-energy matrices $\Sigma_1(E)$ and $\Sigma_2(E)$, the procedure is mechanical: Eqs.(9.1.1) and (9.1.2) can be applied blindly to evaluate the transmission. Of course, complicated contacts can require some extra effort to evaluate the appropriate self-energy matrices, but it is a one-time effort. Besides, as we have mentioned earlier, one can make a reasonable guess based on Eqs.(8.3.12) and (8.3.14) without a detailed calculation – a procedure that can be justified physically by arguing that one never knows the precise shape of the contacts anyway. The examples we discuss below are all based on one-dimensional leads for which the self-energy is written down easily.

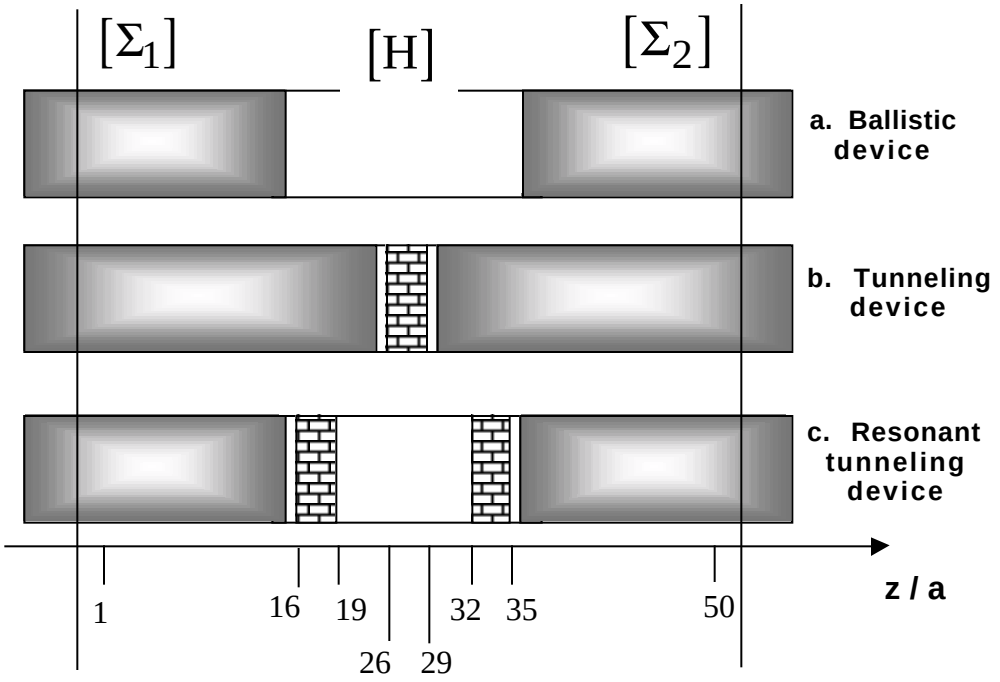


Fig.9.5.1. Three device examples: (a) Ballistic device, (b) Tunneling device and (c) Resonant tunneling device. Barrier regions denoted by ▭ have a conduction band edge 0.4 eV higher.

We use a ***one-dimensional*** discrete lattice with a = 0.3 nm to model each of the following devices which are assumed to be single-moded in the transverse (x- and y-) directions (Fig.9.5.1). The barrier regions indicated as ▭ have a conduction band that is 0.4 eV higher than the rest. We assume that the effective mass (m_c= 0.25m) is the same

everywhere. Fig.9.5.2 shows the (non self-consistent) equilibrium band diagram and transmission functions $\overline{T}(E)$ calculated numerically for each of these devices from the Hamiltonian matrix [H] and the self-energy matrices $\Sigma_{1,2}(E)$.

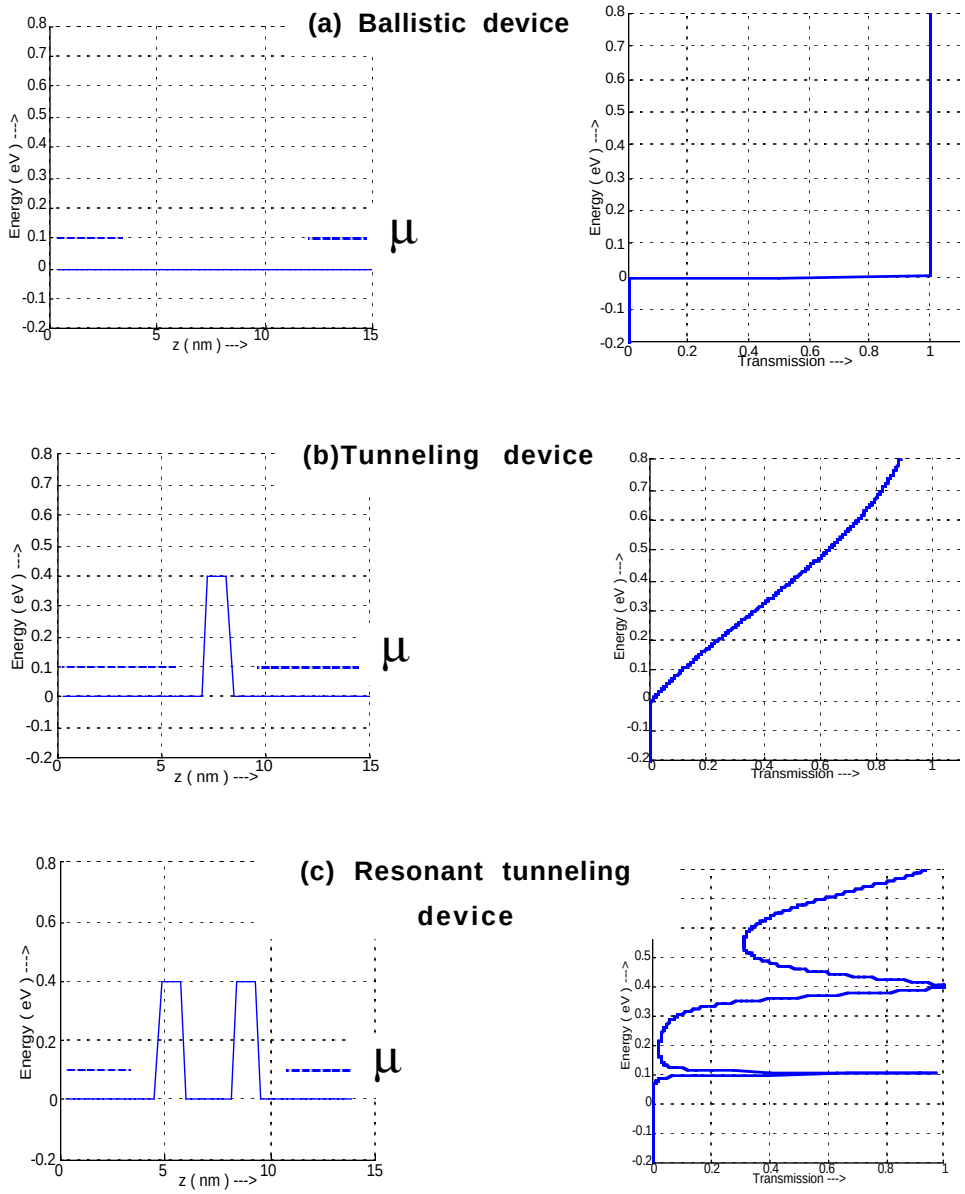
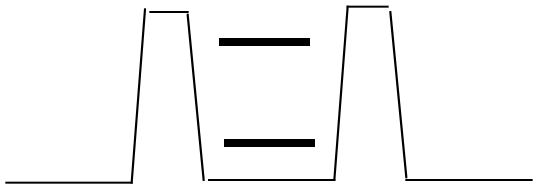


Fig.9.5.2. Equilibrium band diagram and transmission function for each of the devices in Fig.9.5.1.

For the ballistic device the transmission is zero for energies below the band-edge E_c and increases to one above the band-edge. For the tunneling device, the transmission increases from zero to one, though more slowly. The transmission for a resonant tunneling

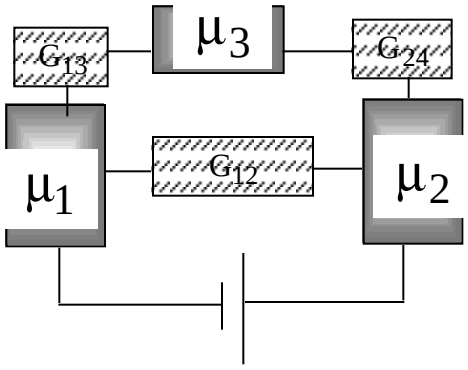
Fig. 9.5.3.



device, on the other hand, shows a very different behavior with two sharp resonances that can be understood by noting that the two barriers create a "box" with discrete energy levels (see Chapter 1, Section 1.1). The transmission from left to right peaks whenever the energy matches one of these levels. It is possible to obtain the same results by matching wavefunctions and derivatives across different sections, but the process quickly gets cumbersome. Arbitrary potential profiles, however, are easily included in the Hamiltonian [H] and the transmission is then calculated readily from the Green's function formalism: $\overline{T}(E) = \text{Trace}\,[\Gamma_1 G \Gamma_2 G^+]$.

In calculating the transmission through devices with sharp resonances (like the resonant tunneling device) it is often convenient to include a Buttiker probe (see Section 9.4) The reason is that it is easy to miss very sharp resonances in a numerical calculation if the energy grid is not fine enough. A Buttiker probe simulates the role of phase-breaking

Fig. 9.5.4.



processes thereby broadening the resonance. The effective transmission is calculated by solving the Buttiker equations (see Eq.(9.4.5)) as explained in the last Section. In this case the transmission between different terminals is reciprocal so that we can calculate the effective transmission from a simple resistor network (see Fig. 9.4.3) adapted to three terminals.

Noting that the conductance is proportional to the transmission we can write the effective transmission using the elementary law of addition for conductors in series and in parallel:

$$\overline{T}_{eff}(E) = \overline{T}_{12}(E) + \frac{\overline{T}_{13}(E)\,\overline{T}_{23}(E)}{\overline{T}_{13}(E) + \overline{T}_{23}(E)} \qquad (9.5.7)$$

Fig.9.5.5 shows the effective transmission for a resonant tunneling device with one Buttiker probe attached to the center of the device. Compared to the earlier result without a probe, the resonances are broadened somewhat, especially the sharpest one.

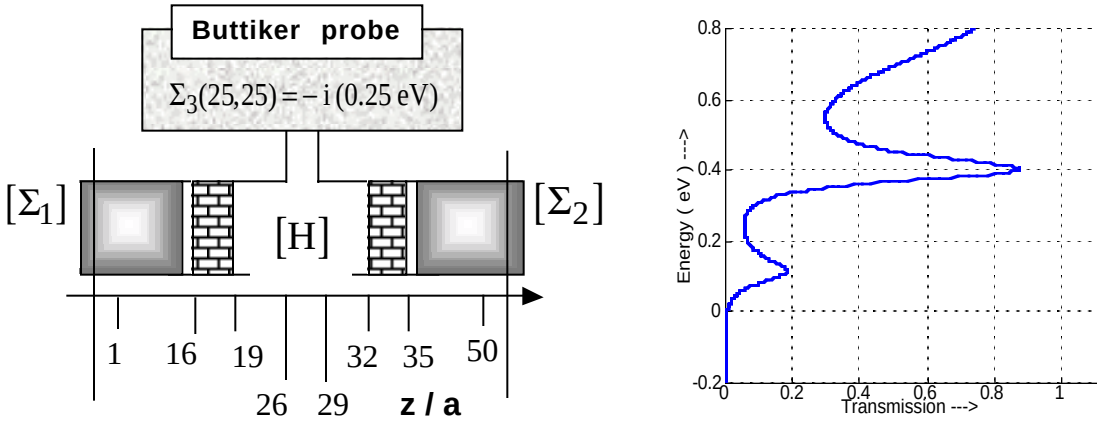
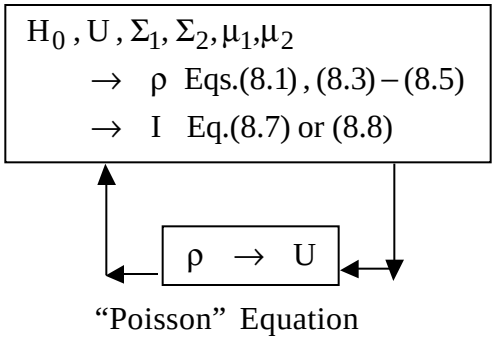


Fig.9.5.5. Effective transmission function for a resonant tunneling device including a Buttiker probe located at lattice site # 25 at the center of the device to simulate the effect of phase-breaking processes phenomenologically. Dotted curve shows result from Fig.9.4.2c without a Buttiker probe.

Current (I) – Voltage (V) characteristics: Eq.(9.1.9) can be used to calculate the I-V characteristics of any coherent device, provided we know how the applied voltage drops

across the device. This is not important if we are only interested in the low bias conductance (or "linear response"), but can be of paramount importance in determining the shape of the full current-voltage characteristics as discussed in Section 1.4.

Fig. 9.5.6.



In general, for quantitatively correct results, it is important to solve for the potential profile ***self-consistently.*** Just like the equilibrium problem (see Fig.7.2.1), we should include a self-consistently determined potential U in the total Hamiltonian $H = H_0 + U([\delta\rho])$.

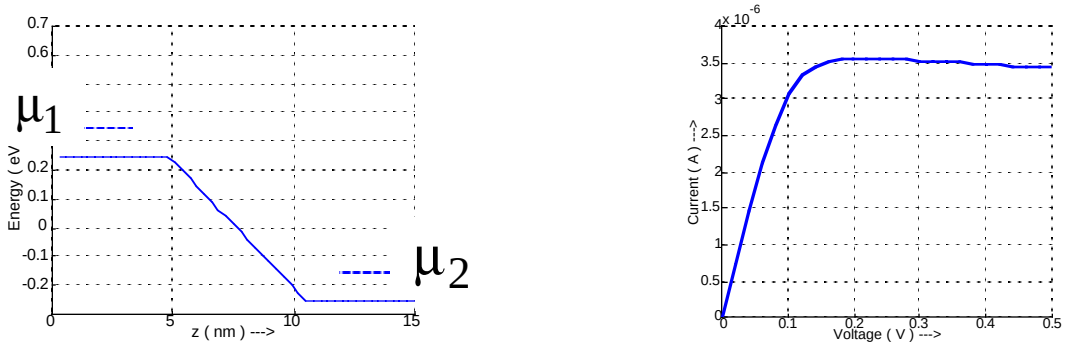
This potential U represents the average potential that an electron feels due to the change $\delta\rho$ in the electron density, or more generally the density matrix. The first step in this process is to calculate the electron density from the diagonal elements of the density matrix. This electron density can then be used in the Poisson equation to calculate the potential which is then included in the Hamiltonian to recalculate the electron density and so on till the process converges as sketched above. A full self-consistent calculation like this can be time-consuming (we will describe a simple one in Section 11.4) and so it is common to assume a "reasonable" potential profile. What is a reasonable profile?

The basic principle is straightforward. If the channel were insulating (low quantum capacitance, see Eq.(7.3.8)), the potential profile would be given by the Laplace potential $U_L(\vec{r})$, obtained by solving the Laplace equation. But if it were metallic (large quantum capacitance), the profile would be given by the "neutral potential" $U_N(\vec{r})$ obtained from the transport equation assuming perfect space charge neutrality everywhere. The correct potential profile is intermediate between these extremes. In regions of low density of states the quantum capacitance is small and the potential profile will tend to follow $U_L(\vec{r})$ while in regions with high density of states the quantum capacitance is large and the potential profile will tend to follow $U_N(\vec{r})$. The common practice for choosing a "reasonable profile" is to assume that the potential follows $U_N(\vec{r})$ (that needed to maintain charge neutrality) at the

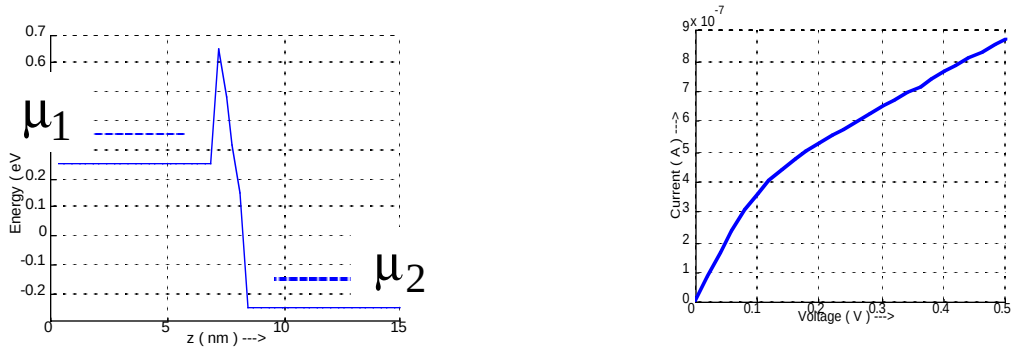
ends which should be regions of high density of states, while in the central channel region the profile is assumed to follow the Laplace potential $U_L(\vec{r})$.

Fig.9.5.7 shows the I-V characteristics for (a) the ballistic device, (b) the tunneling device and (c) the resonant tunneling device calculated assuming that the potential drops linearly across the central unshaded region in Fig.9.5.1. This assumed potential profile gives reasonable qualitative features, but it is easy to check that the results can change quantitatively if we choose different profiles. We will talk about this further in Section 11.4 when we discuss the factors that influence the ON current of a nanotransistor.

(a) Ballistic device



(b)Tunneling device



(c) Resonant tunneling device

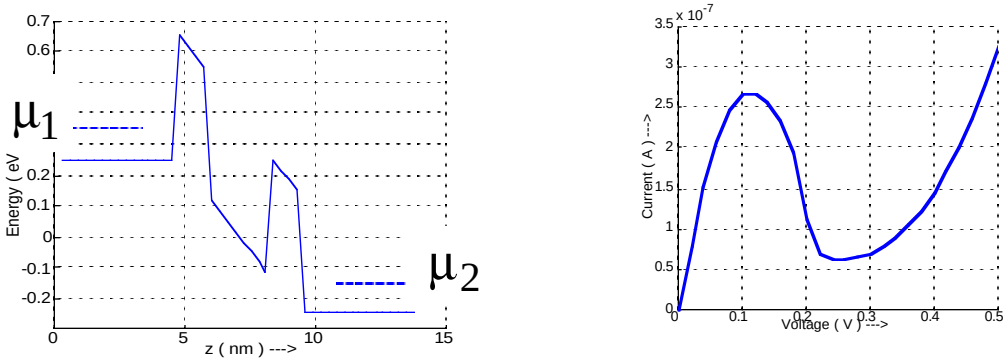


Fig.9.5.7. Current (I) versus voltage (V) characteristics of the three devices shown in Fig.9.5.2 calculated assuming the linear potential profile shown. The left panel shows the assumed band diagram at a bias of 0.5 V.

Exercises

E.9.1. Use a ***one-dimensional*** discrete lattice with a = 0.3 nm to model each of the devices shown in Fig.9.3.1 which are assumed to be single-moded in the transverse (x- and y-directions). Assume that the effective mass (m_c = 0.25m) is the same everywhere. The barrier regions indicated as ▭ have a conduction band that is 0.4 eV higher than the rest.

(a) Set up an energy grid over the range -.2 eV < E < .8 eV and plot the transmission probability as a function of energy. Compare with ***Fig.9.3.2***.

(b) Plot the transmission probability as a function of energy for the ***resonant tunneling device*** using a Buttiker probe as indicated in ***Fig.9.3.3***.

E.9.2. (a) Calculate the current (I) – voltage (V) characteristics in the bias range of 0 < V < 1V. assuming that the applied bias drops across the device following the profile shown in ***Fig.9.4.1.*** Assume the equilibrium Fermi energy to be $E_f = 0.1\,eV$ and the chemical potentials in the two contacts under bias to be $\mu_1 = E_f + qV/2$ and $\mu_2 = E_f - qV/2$. The energy integration needs to be carried out only over the range $\mu_1 + 4k_BT < E < \mu_2 - 4k_BT$. Use an energy grid with $\Delta E \approx 0.2\,k_BT$.

(b) Calculate the electron density n(x) per unit length assuming that the applied bias of 0.5V drops across the ***tunneling device*** following the profile shown in Fig.9.4.1.

E.9.3. ***Transfer Hamiltonian****:* Starting from the expression for the transmission in Eq.(9.3.2), $\overline{T}(E) = \text{Trace}\,[\Gamma_1 G \Gamma_2 G^+]$

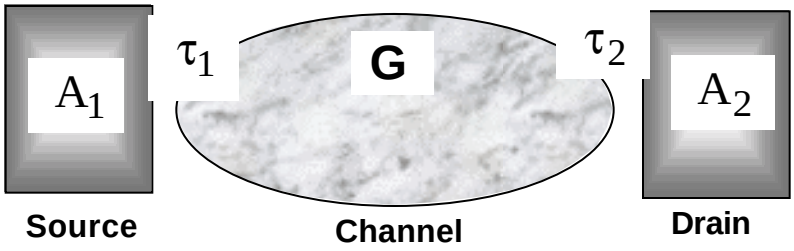


Fig. E.9.3

and making use of the expressions for the broadening matrices in Eq.(9.1.22) show that

$$\overline{T}(E) = \text{Trace}\,[A_1 M A_2 M^+]$$

where A_1 and A_2 are the spectral functions in the two contacts and the matrix element M is given by $M = \tau_1^+ G \tau_2$.

This form is similar to the version often seen in connection with the transfer Hamiltonian formalism (see for example, Eq.(2.3.5) on p.69 of C.J. Chen, Introduction to Scanning Tunneling Microscopy, Oxford (1993)). In the transfer Hamiltonian formalism the matrix element M is assumed to be unaffected by the coupling to the contacts which is assumed to be small, But in the present formulation 'G' and hence 'M' is affected by the contacts through the self-energy due to the contacts.

E.9.4. ***2-D cross-section:*** In the examples of Sections 9.3, 9.4 we have assumed that the device is one-dimensional. The 2-D cross-section can be included in a simple way, if we assume periodic boundary conditions and assume that all the transverse modes are decoupled as we did when calculating the capacitance in Chapter 6. We could then simply sum our 1-D result over all the transverse modes represented by the two-dimensional vector $\vec{k}$ to write ($\varepsilon_{\vec{k}} = \hbar^2 k^2 / 2m_c$):

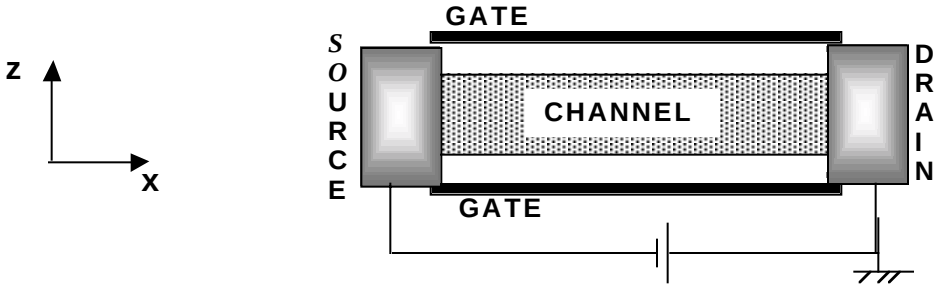
$$I = \frac{q}{2\pi\hbar} \sum_{\vec{k}} \int_{-\infty}^{+\infty} dE\, \bar{T}(E) \left[f_0(E + \varepsilon_{\vec{k}} - \mu_1) - f_0(E + \varepsilon_{\vec{k}} - \mu_2) \right]$$

The transmission function depends only on the longitudinal energy E while the Fermi functions are determined by the total energy $E + \varepsilon_{\vec{k}}$. The summation over $\vec{k}$ can be carried out analytically to write (S: cross-sectional area)

$$\frac{I}{S} = \frac{q}{\pi\hbar} \int_{-\infty}^{+\infty} dE\, \bar{T}(E) \left[f_{2D}(E - \mu_1) - f_{2D}(E - \mu_2) \right]$$

This means that the current in a device with a 2-D cross-section is obtained using the same procedure that we used for a 1-D device, provided we use the k-summed Fermi function f_{2D}(see Eq.(6.2.12) in place of the usual Fermi function. Repeat Prob.9.2 using f_{2D} (see Eq.(6.2.12)) instead of the Fermi function f_0 to account for a device with a 2-D cross-section. The current should now be expressed in A / m^2 and the electron density should be expressed in / m^3.

E.9.5. 1-D cross-section*:*



In analyzing Field Effect Transistors, we often have a 1-D cross-section (y-direction) to sum over, while the transmission has to be calculated from a 2-D problem in the z-x plane. Assuming periodic boundary conditions in the y-direction show that the 1-dimensional k-sum can be done analytically to obtain

$$\frac{I}{W} = \frac{q}{\pi\hbar}\int_{-\infty}^{+\infty} dE\ \bar{T}(E)\left[f_{1D}(E-\mu_1)-f_{1D}(E-\mu_2)\right]$$

where the 1-D k-summed Fermi function is given by

$$f_{1D}(E) \equiv \left(\frac{m_c k_B T}{2\pi\hbar^2}\right)\Im_{-1/2}\left(-\frac{E}{k_B T}\right)$$

with $$\Im_{-1/2}(x) \equiv \frac{1}{\sqrt{\pi}}\int_0^{+\infty}\frac{dy}{1+\exp(y-x)}\frac{1}{\sqrt{x}} = \frac{d}{dx}\Im_{+1/2}(x)$$

where $\Im_{+1/2}(x)$ was defined in Eq.(7.2.23).