Lecture 23

Does NEGF Include “Everything?”

23.1. Coulomb Blockade
23.2. Fock Space Description
23.3. Entangled States

Back in Lecture 18 we used this picture to introduce our quantum transport model representing an elastic channel described by a Hamiltonian $[H]$ and self-energies $[\Sigma]$ describing the exchange of electrons with the physical contacts and energy with the surroundings which can be viewed as additional conceptual “contacts”.

Given these inputs, the basic NEGF equations (see Eqs.(19.1)-(19.4)) tell us how to analyze any given structure. Since then we have been looking at various examples illustrating how one writes down $[H]$ and $[\Sigma]$ and uses the NEGF equations to extract concrete results and investigate the physics. One major simplification we have adopted is in our treatment of the interactions in the channel represented by $\Sigma_0$ which we have either ignored (coherent transport) or treated as an elastic dephasing process described by Eqs.(19.35).
This choice of self-energy functions leads to no exchange of energy with the surroundings, but it has an effect on transport due to the exchange of momentum and "phase". Basically we have been talking about elastic resistors like the ones we started these Lectures with, except that we are now including quantum mechanical effects. One could say that in the last few Lectures we have applied the general Non-Equilibrium Green’s Function (NEGF) method to an elastic resistor, just as in Part one we applied the general Boltzmann Transport Equation (BTE) to an elastic resistor.

So how do we go beyond elastic resistors? For semiclassical transport, it is clear in principle how to include different types of interaction into the BTE for realistic devices and much progress has been made in this direction. Similarly for quantum transport, the NEGF tells us how to evaluate the self-energy $\Sigma_0$ for any given microscopic interaction. In these lectures we have talked only about elastic dephasing which is a small subset of the interactions considered in the classic work on NEGF (see for example, Danielewicz 1984 or Mahan 1987).

In practice, however, it remains numerically challenging to go beyond elastic resistors and approximate methods continue to be used widely. Readers interested in the details of device analysis at high bias may find an old article (Datta (2000)) useful. This article has a number of concrete results obtained using MATLAB codes that I had offered to share with anyone who asked me for it. Over the years many have requested these codes from me which makes me think they may be somewhat useful and we plan to have these available on our website for these notes.

I should mention that many devices are rather forgiving when it comes to modeling the physics of inelastic scattering correctly. Devices with energy levels that are equally connected to both contacts (Fig.9.5b.) do not really test the deep physics of inelastic transport and cannot distinguish between a good theory and a bad one. A good test for inelastic scattering models is the device shown in Fig.9.5a for which the entire terminal current is driven by inelastic processes. Only a fundamentally sound theory will predict results that comply with the requirements of the second law.

But practical issues apart, can the NEGF method model “everything”, at least in principle?

The formal NEGF method developed in the 1960’s was based on many-body perturbation theory (MBPT) which provided clear prescriptions for evaluating the self-energy functions

$$\Sigma, \Sigma^{in}$$

for a given microscopic interaction up to any order in perturbation theory. It may seem that using MBPT we can in principle include everything. However, I believe this is not quite true since it is basically a perturbation theory which in a broad sense amounts to evaluating a quantity like $(1-x)^i$ by summing a series like $1+x+x^2+x^3+...$, which works very well if $x$ is much less than one. But if $x$ happens to exceed one, it does not work and one needs non-perturbative methods, or perhaps a different perturbation parameter.

This is one of the reasons I prefer to decouple the NEGF equations (Eqs.(19.1) through (19.4)) from the MBPT-based methods used to evaluate the self-energy functions. The latter may well evolve and get supplemented as people find better approximations that capture the physics in specific situations.

With equilibrium problems, for example, density functional theory (DFT)-based techniques have proven to be very successful and are often used in quantum chemistry in place of MBPT. I believe one should be cautious about expecting the same success with non-equilibrium problems where a far greater spectrum of many body states are made accessible and can be manipulated through a judicious choice of contacts, but it is quite likely that people will find insightful approaches that capture the essential physics in specific problems.
Like the BTE for semiclassical transport, NEGF-based methods in their simplest form, seem to provide a good description of problems where electron-electron interactions can be treated within a mean field theory based on the widely used picture of quasi-independent electrons moving in a self-consistent potential \( U \) due to the other electrons (Section 18.2).

As we saw in Lecture 8, for low bias calculations one needs to consider only the equilibrium potential which is already included in the semi-empirical tight-binding (TB) parameters used to construct our Hamiltonian \( [H] \). For real devices operating at high bias, the change in the potential due to any changes in the electron occupation in the channel are routinely included using the Poisson equation which is the simplest approximation to the very difficult problem of electron-electron interactions and there have been extensive discussions of how the self-consistent field (scf) can be corrected to obtain better agreement with experimental results.

However, there are examples where the self-consistent field approach itself seems to fail and some of the most intriguing properties arise from a failure of this simple picture. The purpose of this Lecture is to alert the reader that a straightforward application of NEGF may well miss these important experimentally observable effects. Future challenges and opportunities may well involve effects of this type, requiring insightful choices for \( \Sigma \), \( \Sigma'^{\infty} \) if we wish to use the NEGF method.

23.1. Coulomb Blockade

In the spirit of the bottom-up approach let us consider the simplest resistor that will show this effect, one that is only slightly more complicated than the one-level resistor we started with (Fig.3.1). We assume two levels, a spin up and a spin down, having the same energy \( \varepsilon \), with the equilibrium chemical potential \( \mu \) located right at \( \varepsilon \), so that each level is half-filled since the Fermi function \( f_{\varepsilon}(E=\mu) \) equals 0.5. Based on what we have discussed so far we would expect a high conductance since the electrochemical potential lies right in the middle of each broadened level as shown in the upper sketch in Fig.23.1.

However, if the single electron charging energy \( U_0 \) is large then the picture could change to the lower one where one level floats up by \( U_0 \) due to the electron occupying the other level. Why doesn't the other level float up as well? Because no level feels any potential due to itself. This self-interaction correction is missed in the self-consistent field (SCF) model discussed in Lecture 8 where we wrote \( U = U_0 N \). Instead we need an unrestricted SCF where each level \( i \) is not restricted to feeling the same potential. Instead it feels a potential \( U_i \) that depends on the change in the number of electrons occupying all levels except for \( i \):

\[
U_i = U_0(N - N_i)
\]  

If we were to use Eq.(23.1) instead of \( U = U_0 N \) we would obtain a picture like the lower one in Fig.25.4, assuming that \( \mu \) is adjusted to have approximately one electron inside the channel. We would find a self-consistent solution with

\[
N_{dn} = 1, U_{up} = 0, N_{up} = 0, U_{dn} = 0
\]
The down level will be occupied \((N_{dn} = 1)\) and the resulting potential \((U_{up} = U_0)\) will cause the up level to float up and be unoccupied \((N_{up} = 0)\). Because it is unoccupied, the potential felt by the down level is zero \((U_{dn} = 0)\), so that it does not float up, consistent with what we assumed to start with.

Of course, the solution with up and down interchanged

\[ N_{up} = 1, U_{dn} = 0, N_{dn} = 0, U_{up} = 0 \]

is also an equally valid solution. Numerically we will converge to one or the other depending on whether we start with an initial guess that has more \(N_{up}\) or \(N_{dn}\). Experimentally the system will fluctuate between the two solutions randomly over time.

Why have we not worried about this before? Because it is not observable unless the charging energy \(U_0\) is well in excess of both \(kT\) and the broadening. \(U_0/q\) is the potential the channel would float to if one electron were added to it. For a large conductor this potential is microvolts or smaller and is unobservable even at the lowest of temperatures. After all, any feature in energy is spread out by \(kT\) which is \(-25\) meV at room temperature and \(-200\) μeV at \(-1\)K. The single electron charging effect that we are talking about, becomes observable at least at low temperatures, once the conductor is small enough to make \(U_0\) of the order of a meV. For molecular sized conductors, \(U_0\) can be hundreds of meV making these effects observable even at room temperature.

Fig. 23.2. The single electron charging energy \(U_0\) is the electrostatic energy associated with one extra electron in the channel.

However, there is a second factor that also limits the observability of this effect. We saw in Lecture 19 that in addition to the temperature broadening \(-kT\), there is a second and more fundamental broadening, \(\gamma \sim h/T\) related to the transfer time. Single electron charging effects will be observed only if the Coulomb gap \(U_0\) exceeds this broadening: \(U_0 >> h/T\). For this reason we would not expect to see this effect even in the smallest conductors, as long as it has good contacts.

### 23.1.1 Current versus voltage

Let us now move on from the low bias conductance to the full current-voltage characteristics of the two-level resistor. For simplicity we will assume that the levels remain fixed with respect to the source and are unaffected by the drain voltage, so that we do not have to worry about the kind of issues related to simple electrostatics that we discussed in Lecture 8.

A simple treatment ignoring electron-electron interactions then gives the curve marked "non-interacting" in Fig. 23.3. Once the electrochemical potential \(\mu_2\) crosses the levels at \(\varepsilon\), the current steps up to its maximum value.

If we now include charging effects through a self-consistent potential \(U = U_0\), the current step stretches out over a voltage range of \(-U_0/q\), since the charging of the levels makes them float up and it takes more voltage to cross them completely.

Fig. 23.3. Current-voltage characteristic of a two-level resistor with \(U = 0\) and with \(U = U_0 N\)
But if we include an SCF with self-interaction correction (Eq. (23.1)) we calculate a current-voltage characteristic with an intermediate plateau as shown in Fig. 23.4 which can be understood in terms of the energy level diagrams shown. At first only the lower level conducts giving only half the maximum current and only when the voltage is large enough for \( \mu_2 \) to cross \( \varepsilon + U_0 \) that we get the full current.

Such intermediate plateaus in the I-V characteristics have indeed been observed but the details are not quite right. The correct plateau current is believed to be \( 2/3 \) and not \( 1/2 \) of the total current of \( 2q/I \). This represents an effect that is difficult to capture within a one-electron picture, though it can be understood clearly if we adopt a different approach altogether, which we will now describe.

![Fig. 23.4. Current-voltage characteristic of a two-level resistor.](image)

23.2. Fock Space Description

This approach is based on the Fock space picture introduced in Lecture 16. As we discussed earlier, in this new picture we do not think in terms of one-electron levels that get filled or emptied from the contacts. Instead we think in terms of the system being driven from one state to another.

![One-electron picture versus Fock space picture for a one-level channel.](image)

For example Fig. 23.5 shows how we would view the one-level resistor in this Fock space picture where the system can be one of two states: \( 0 \) representing an empty state, and \( 1 \) representing a full state. Fig. 23.6 shows the two pictures for a two-level resistor. In general a \( N \)-level resistor will have \( 2^N \) Fock space states.

![One-electron picture versus Fock space picture for a two-level channel.](image)

23.2.1. Equilibrium in Fock space

As we discussed in Lecture 16, there is a well-defined procedure for finding the probabilities of finding the system in a given eigenstate \( i \) at equilibrium.
\[ p_i = \frac{1}{Z} e^{-(\epsilon_i - \mu N_i)/kT} \]  
(Same as Eq.(16.14))

We could use this to calculate any equilibrium property. For example, suppose we want to find the number of electrons, \( n \), occupying the two-level channel shown in Fig.23.6 if it is in equilibrium with an electrochemical potential \( \mu \).

Fig.23.7 shows the result obtained by plotting \( n \) versus \( \mu \) from the equation

\[ n = \sum_i N_i p_i = p_{01} + p_{10} + 2p_{11} \]

using the equilibrium probabilities from Eq.(16.14) cited above. Note how the electron number changes by one as \( \mu \) crosses \( \epsilon \) and then again when \( \mu \) crosses \( \epsilon + U_0 \) in keeping with the lower picture in Fig.23.1.

Note, however, that we did not assume the picture from Fig.23.1 with two one-electron states at different energies. We assumed two one-electron states with the same energy (Fig.23.6) but having an interaction energy that is included in the Fock space picture.

If we are interested in the low bias conductance \( G \) as a function of \( \mu \), we could deduce it from the \( n(\mu) \) plot in Fig.23.7. As we discussed in Lecture 2, current flow is essentially because the two contacts with different \( \mu \)'s have different agendas, since one likes to see more electrons in the channel than the other. From this point of view one could argue that the conductance should be proportional to \( dn/d\mu \) and show peaks at

\[ \mu = \epsilon \quad \text{and at} \quad \mu = \epsilon + U_0 \]

as shown. This is indeed what has been observed experimentally for the low bias conductance of small conductors in the single-electron charging regime where \( U_0 \) exceeds both the thermal energy \( kT \) and the energy broadening due to contacts.

As we saw in Lecture 15, low bias conductance is an equilibrium property that can be deduced using the principles of equilibrium statistical mechanics. Current flow at higher voltages on the other hand requires the methods of non-equilibrium statistical mechanics. Let me explain briefly how one could understand the 2/3 plateau shown in Fig.23.4 by calculating the current at high bias in the Fock space picture.

### 23.2.2. Current in the Fock space picture

To calculate the current we write an equation for the probability that the system will be found in one of its available states, which must add up to one. For example for the one level resistor we could write

\[ v_1 p_0 = v_2 p_1 \rightarrow \frac{p_1}{p_0} = \frac{v_1}{v_2} \rightarrow p_1 = \frac{v_1}{v_1 + v_2} \]
assuming that the left contact sends the system from the 0 state to the 1 state at a rate $V_1$, while the right contact takes it in the reverse direction at a rate $V_2$ and at steady-state the two must balance. The current is given by

$$I = qV_2 p_1 = q \frac{V_1 V_2}{V_1 + V_2}$$

(23.2)

in agreement with our earlier result in Lecture 19 (see Eq.(19.10b)) obtained from a one-electron picture.

But the real power of this approach is evident when we consider levels with multiple interacting levels. Consider for example the two-level resistor biased such that electrons can come in from the left contact and transfer the system from 00 to 01 or to 10, but not to the 11 state because of the high charging energy $U_0$. This is the biasing condition that leads to a plateau at 2/3 the maximum value (Fig.23.4) that we mentioned earlier.

In this biasing condition, the system can only come out of the 11 state, but never transfer into it, and so the steady-state condition can be calculated simply by considering the kinetics of the three remaining states in Fock space, namely 00, 01 and 10:

$$2V_1 p_{00} = V_2 (p_{01} + p_{10})$$

$$\rightarrow p_{01} + p_{10} = \frac{2V_1}{V_2}$$

$$\rightarrow p_{01} + p_{10} = \frac{2V_1}{2V_1 + V_2}$$

where we have made use of the requirement that all three probabilities must add up to one. Hence

$$I = qV_2 (p_{01} + p_{10}) = q \frac{2V_1 V_2}{2V_1 + V_2}$$

With $V_1 = V_2 \rightarrow I = \frac{2}{3} qV_1$

which is 2/3 the maximum current as stated earlier.

It is important to note the very special nature of the solution we just obtained which makes it hard to picture within a one-electron picture. We showed that the system is equally likely to be in the states 00, 01 and the 10 states, but zero probability of being in the 11 state.

Fig.23.8
The intermediate plateau in the current corresponds to the channel being in a strongly correlated state.

In other words, if we looked at the up-spin or the down-spin state (in the one-electron picture) we would find them occupied with 1/3 probability. If electrons were independent then we would expect the probability for both to be occupied to be the product = 1/9.

Instead it is zero, showing that the electrons are correlated and cannot be described with a one-electron occupation factor $f$ of the type we have been using throughout these lectures. Even with quantum transport we replaced the $f$'s with a matrix $G'$ obtained by summing the $\Psi \Psi^+$ for individual electrons. This adds sophistication to our understanding of the one-electron state, but it still does not tell us anything about two-electron correlations.
23.3. Entangled states

What we just saw with one quantum dot is actually just the proverbial tip of the iceberg. Things get more interesting if we consider two or more quantum dots.

For example, with two coupled quantum dots we could write the one-electron Hamiltonian matrix as a 4x4 matrix using the up and down states in dots 1 and 2 as the basis functions as follows:

\[
[H] =
\begin{array}{cccc}
    u_1 & u_2 & d_1 & d_2 \\
    u_1 & \varepsilon_1 & t & 0 & 0 \\
    u_2 & t & \varepsilon_2 & 0 & 0 \\
    d_1 & 0 & 0 & \varepsilon_1 & t \\
    d_2 & 0 & 0 & t & \varepsilon_2 \\
\end{array}
\]  

(23.3)

If there were no inter-dot coupling then these sixteen states would be the eigenstates and we could analyze their dynamics in Fock space just as we did for one dot. But in the presence of inter-dot coupling the true eigenstates are linear combinations of these states and these entangled states can lead to novel physics and make it much more interesting.

The 0-electron and 4-electron states are trivially composed of just one Fock space state, while the 1-electron state is essentially the same as the states in a one-electron picture. Indeed the 3-electron state also has a structure similar to the one-electron state and could be viewed as a 1-hole state.

The 2-electron states, however, have an interesting non-trivial structure. Consider the six 2-electron states which we label in terms of the two states that are occupied: \(u_1d_1, u_2d_2, u_1d_2, u_2d_1, u_1u_2, d_1d_2\). Using these we can write the Fock space Hamiltonian \([HH]\) as explained below.

The diagonal elements of \([HH]\) are written straightforwardly by adding the one-electron energies plus an interaction energy \(U_0\) if the two basis functions happen to be on the same dot making their Coulomb repulsion much stronger than what it is for two states on neighboring dots.
\[
[H]H = \begin{bmatrix}
    u_d^1 & u_d^2 & u_d^1 & u_d^2 & u_u^1 & u_u^2 & d_d^1 & d_d^2 \\
    u_d^1 & 2e_1 + U_0 & t & t & 0 & 0 & 0 & 0 \\
    u_d^2 & 0 & 2e_2 + U_0 & t & t & 0 & 0 & 0 \\
    u_d^1 & t & t & e_1 + e_2 & 0 & 0 & 0 & 0 \\
    u_d^2 & t & t & 0 & e_1 + e_2 & 0 & 0 & 0 \\
    u_u^1 & 0 & 0 & 0 & e_1 + e_2 & 0 & 0 & 0 \\
    u_u^2 & 0 & 0 & 0 & 0 & e_1 + e_2 & 0 & 0 \\
    d_d^2 & 0 & 0 & 0 & 0 & 0 & e_1 + e_2 & 0 
\end{bmatrix}
\] (23.4)

The off-diagonal entries \( t \) are obtained by noting that this quantity couples the one electron states \( u_1 \) to \( u_2 \) and \( d_1 \) to \( d_2 \). With two electron states we have inserted \( t \) for non-diagonal elements that couples those states for which one state remains unchanged while the other changes from \( u_1 \) to \( u_2 \) or from \( d_1 \) to \( d_2 \).

The lowest eigenstate obtained from the two-electron Hamiltonian in Eq.(23.4) is with a wavefunction of the form \( (s_1, s_2 < 1) \)

\[
S: \, S \{ \{u_d^1\} + \{u_d^2\}\} + s_1 \{u_d^1\} + s_2 \{u_d^2\} \] (23.5)

is called the singlet state. Next comes a set of three states (called the triplets) that are higher in energy. These have the form

\[
T1: \, \frac{1}{\sqrt{2}} \{ \{u_d^1\} - \{u_d^2\}\} \\
T2: \, \{u_u^1\} \\
T3: \, \{d_d^1\} 
\] (23.6)

A system with two electrons is normally viewed as occupying two one-electron states. The states \( T2, T3 \) permit such a simple visualization. But the states \( S \) and \( T1 \) do not.

For example, each term in the state

\[
T1: \, \frac{1}{\sqrt{2}} \{ \{u_d^1\} - \{u_d^2\}\} 
\]

permits a simple visualization: \( \{u_d^1\} \) stands for an upspin electron in 1 and a downspin electron in 2 while \( \{u_d^2\} \) represents an upspin in 2 and a downspin in 1. But the real state is a superposition of these two "simple" or unentangled states and there is no way to define two one-electron states \( a \) and \( b \) such that the two-electron state could be viewed as \( \{ab\} \). Such states are called entangled states which comprise the key entity in the emerging new field of quantum information and computing.

How would we compute the properties of such systems? The equilibrium properties are still described by the general law of equilibrium stated earlier

\[
p_i = \frac{1}{Z} e^{-(E_i - \mu N_i)/kT} \quad \text{(Same as Eq.(16.14))}
\]

and using the equilibrium properties to evaluate the average number of electrons.

\[
n = \sum_i N_i p_i
\]

The energies \( E_i \) are obtained by diagonalizing the Fock space Hamiltonian \([HH]\) that we just discussed. Fig.23.9 shows the plot of \( n \) versus \( \mu \) which looks like Fig.23.7, but the middle plateau now involves the entangled singlet state just discussed. There is also some additional structure that we will not get into. The main point we wanted to make is that the law of equilibrium statistical mechanics is quite general and can be used in this case.
Fig. 23.9. Equilibrium number of electrons, \( n \) in the two-level channel shown in Fig.23.6 as a function of \( \mu \), assuming \( \epsilon = 10kT \), \( U_0 = 20kT \).

But the calculation of current at high bias is a non-equilibrium problem that is not as straightforward. Using the entangled states one could set up a rate equation as we did in the last Section and understand some of the interesting effects that have been observed experimentally including negative differential resistance (NDR), that is a decrease in current with increasing voltage (see for example Muralidharan et al. 2007). More generally one needs quantum rate equations to go beyond the simple rate equations we discussed and handle coherences (Braun et al. 2004, Braun and Brouwer 2005).

Can we model transport involving correlated and/or entangled states exactly if we use a Fock space picture instead of using NEGF and including interactions only approximately through self-energies? Sort of, but not quite.

There are two problems. The first is practical. A \( N \)-level problem in the one-electron picture escalates into a \( 2^N \) level problem in the Fock space picture. The second is conceptual.

We saw in Lecture 19 how the NEGF method allows us to include quantum broadening in the one-electron Schrödinger equation. To our knowledge there is no comparable accepted method for including broadening in the Fock space picture. So the rate equation approach from the last Section works fine for weakly coupled contacts where the resulting broadening is negligible, but the regime with broadening comparable to the charging energy stands out as a major challenge in transport theory. Even the system with two levels (Fig.23.7) shows interesting structure in \( n(\mu) \) in this regime ("Kondo peak") that has occupied condensed matter physicists for many decades.

One could view Coulomb blockade as the bottom-up version of the Mott transition, a well-studied phenomenon in condensed matter physics. In a long chain of atoms, the levels \( \epsilon \) and \( \epsilon + U_0 \) (Fig.23.1) will each broaden into a band of width \( ~2t_0 \), \( t_0 / h \) being the rate at which electrons move from one atomic site to the next. These are known as the lower and upper Hubbard bands. If their separation \( U_0 \) exceeds the width \( 2t_0 \) of each band we will have a Mott insulator where the electrochemical potential lies in the middle of the two bands with very low density of states and hence very low conductance. But if \( U_0 \) is small, then the two bands form a single half-filled band with a high density of states at \( E = \mu_0 \) and hence a high conductance.

Needless to say, the full theory of the Hubbard bands is far more complicated than this oversimplified description might imply and it is one of the topics that has occupied condensed matter theorists for over half a century. Since the late 1980's it has acquired an added significance with the discovery of a new class of superconductors operating at relatively high temperatures above 100K, whose mechanism continues to be controversial and hotly debated.

This problem remains one of the outstanding problems of condensed matter theory, but there seems to be general agreement that the essential physics involves a two-dimensional array of quantum dots with an inter-dot coupling that is comparable to the single dot charging energy.