Non-Equilibrium Green's Function (NEGF) Method: A Different Perspective

Supriyo Datta, email: datta@purdue.edu
Electrical Engineering Building, 465 Northwestern Avenue,
West Lafayette, IN 47907, USA

WATCH VIDEO LECTURE https://nanohub.org/resources/22800

INTRODUCTION

The NEGF method was established in the 1960's through the classic work of Keldysh and others [1] using the methods of many-body perturbation theory (MBPT) and this approach is widely used in the literature [2]. By contrast I have introduced a different approach starting with the one-electron Schrödinger equation [3, 4] which is used by many in the nanoelectronics community. In this talk I will try to answer the questions I often get regarding the relation between the two approaches and I thank the organizers of the IWCE for giving me this opportunity.

SCHRÖDINGER TO NEGF

Let me start by quickly outlining our approach following the discussion in [4]. We start from the usual time-independent Schrödinger equation and add two terms to it representing the outflow and inflow from the contact

$$E\{\psi\} = [H]\{\psi\} + \underbrace{[\Sigma]\{\psi\}}_{OUTFLOW} + \underbrace{\{s\}}_{INFLOW} \quad (1)$$

These two terms arise from imposing open boundary conditions on the Schrödinger equation with an incident wave from the contact as shown in Chapters 8,9 of [3b]. Note the similarity of the additional terms here with those appearing in the Langevin equation used to describe Brownian motion.

Using this modified Schrödinger equation (1), the wavefunction can be written as

$$\{\psi\} = [EI - H - \Sigma]^{-1}\{s\} \quad (2)$$

Since the inflow from multiple sources $\{s\}$ are incoherent, one cannot superpose the resulting $\{\psi\}$'s and it is more convenient to work in terms of quantities like (superscript '+' denotes conjugate transpose)

$$[G^n] \sim \{\psi\}\{\psi\}^+, [\Sigma^{in}] \sim \{s\}\{s\}^+$$

which can be superposed. This leads to the NEGF equations as we will now show.

Defining $G^R = [EI - H - \Sigma]^{-1}$ (3)

and $G^A = [G^R]^+$ we can write from (2)

$$\{\psi\} = [G^R]\{s\}$$

so that $\underbrace{\{\psi\}\{\psi\}^+}_{G^n} = [G^R]\underbrace{\{s\}\{s\}^+}_{\Sigma^{in}}[G^A]$

giving us the second NEGF equation

$$G^n = G^R\,\Sigma^{in}\,G^A \quad (4)$$

Though we have changed the notation, writing Σ for Σ^R, G^n *for* $-iG^<$, Σ^{in} *for* $-i\Sigma^<$ Eqs.(3,4) are essentially the same as Eqs.(75-77) in Keldysh (1965), which is one of the seminal founding papers on the NEGF method that obtained these equations using MBPT.

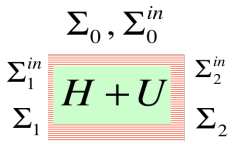
In modern applications, the terminal currents are calculated by supplementing the NEGF equations with the current operator

$$I^{op} = \frac{\Sigma G^n - G^n \Sigma^+}{i\hbar} + \frac{\Sigma^{in} G^A - G^R \Sigma^{in}}{i\hbar} \qquad (5)$$

We will not discuss the current operator further in this talk, but this too can be obtained from the one-electron picture [3a, see Eq.(8.5.7)], though it is commonly obtained using a many-body formalism [2, see Eq.(16.7)]. Also, for simplicity we will only discuss the time-independent version here, a similar derivation could be used for the time-dependent version too [3b, see Appendix].

One Electron or Many-Body?

How could we obtain these results using elementary arguments, without invoking MBPT?



We are considering a conductor with a Hamiltonian H that describes the quasi-particle dynamics around the equilibrium state, along with a suitable self-consistent potential U that approximately accounts for electron-electron interactions. This approach is widely used and the subtle many-body issues associated with the proper choice of U and its limitations have been discussed extensively.

Here, however, we are concerned more about the non-Hermitian parts described by the self-energy and inscattering functions

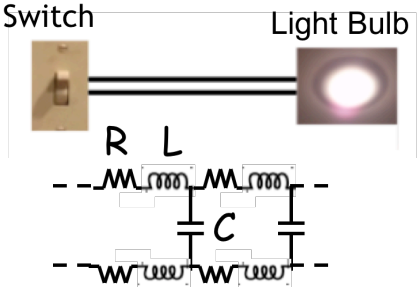
$$\Sigma = \underbrace{\Sigma_1 + \Sigma_2}_{contacts} + \underbrace{\Sigma_0}_{"rigid"} + \underbrace{\left(\Sigma_0\right)_{diss}}_{"many-body"}$$

$$\Sigma^{in} = \underbrace{\Sigma_1^{in} + \Sigma_2^{in}}_{contacts} + \underbrace{\Sigma_0^{in}}_{"rigid"} + \underbrace{\left(\Sigma_0^{in}\right)_{diss}}_{"many-body"}$$

The components labeled 1, 2 arise from the contacts, while the component labeled 0 arises from the interactions of electrons with the surroundings which can be split into two parts: one from rigid scatterers with no internal degrees of freedom, and a many-body part that couples an individual electron into many degrees of freedom. We have labeled this part as "diss" since it usually involves the dissipation of energy, but this is not essential.

It is only this last component that requires many-body theory. The contact components as well as the rigid interaction can be treated straightforwardly within a one-electron picture. Indeed even the "diss" component can be treated within a one-electron picture in the Born approximation, though it requires subtle arguments related to the exclusion principle [3b, Chapter 10].

So why do most serious treatments of transport start from many-body theory? It is probably because electron-electron interaction is generally considered an essential factor without which it is meaningless ***even to start a discussion about resistance.***



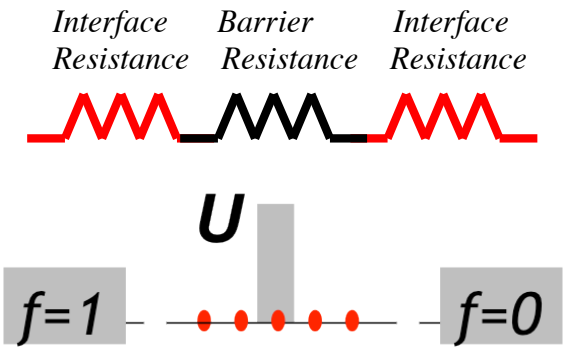
This "gut feeling" comes largely from a fact that we all learn as undergraduates, namely that when we turn on the switch and a light bulb turns on, it is not because an electron actually traveled from the switch to the bulb. That would take far too long. What travels is the signal, as one electron pushes the next one, which in turn pushes the next one and the signal travels close to the velocity of light.

It seems that a model describing current in terms of the flow of individual electrons would predict that signals will travel at the velocity of electrons, which is clearly wrong. Hence it would seem that such a theory cannot be correct even to zeroth order.

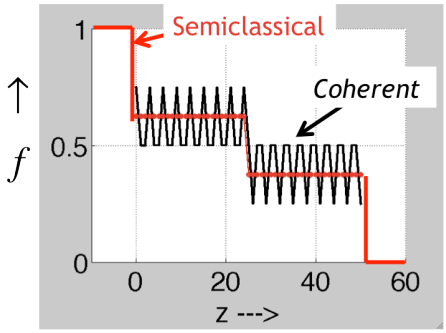
Note, however, that to describe the propagation of transient signals we need a distributed transmission line model that includes not just R, but also an inductance L and a capacitance C as shown above. These quantities could include transport related corrections in small conductors, but are ordinarily determined primarily by magnetostatics and electrostatics [5]. The signal velocity, controlled largely by this L and C, can be well in excess of individual electron velocities, reflecting the collective process described above. However, the low frequency conduction properties are unaffected by L or C, and depend only on the dc resistance R that can usually be understood at least to first order in terms of the transport of individual electrons. Significant many-body corrections may be needed in special cases, but there is nothing fundamentally wrong at least in ***starting the discussion*** without these effects.

An Example

The origin of the resistance R lies in the loss of momentum rather than the loss of energy and this is evident if we plot the occupation of states (which reflects the electrochemical potential) inside a ballistic conductor with a single localized barrier U as shown below.



The potential profile is calculated directly from the NEGF equations [4] using different choices for the interaction term. The first plot is obtained with no interactions in the channel giving an oscillatory profile that can also be obtained from the scattering theory of coherent transport.

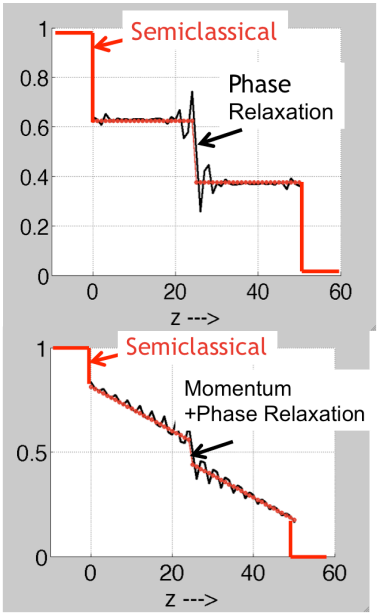


Experimentally these coherent oscillations are seldom observed especially at room temperature [6] due to the incoherent interactions inside the channel. These interactions are difficult to include within the scattering theory of transport, but are straightforwardly included in the NEGF equations through a proper choice of the interaction term in the form

$$\Sigma_0 = D \otimes G^R, \Sigma_0^{in} = D \otimes G^n$$

where through a proper choice of the tensor D we can include different types of elastic interactions. Two such examples are shown below, one including only phase-relaxing interactions, and another including interactions that relax both phase and momentum.

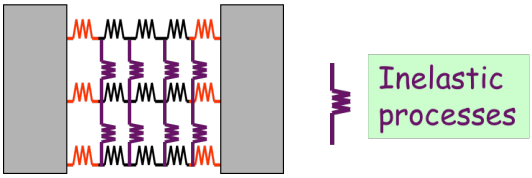
Note that purely phase relaxing mechanisms destroy the coherent oscillations but do not introduce any additional resistance: the potential profile is still flat in the channel except at the barrier. But once we introduce momentum relaxation there is additional resistance in the channel as evidenced by the non-zero slope in the potential [7].



The key point I am trying to make is that the fundamental cause of resistance is the loss of momentum : none of these calculations include any dissipative interactions. The drop in potential across the structure indicates a drop in the occupation of the states in the channel but does not imply any loss of energy or dissipation. The electrons in these regions have a highly non-equilibrium energy distribution and these hot electrons relax their excess energy elsewhere as they flow away, as discussed in Chapter 2 of [3a]. But this loss of energy can occur many microns away from the channel and gives no hint as to what caused the resistance in the first place.

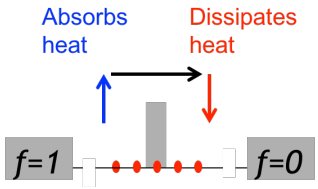
"Vertical" Flow

But can we calculate the resistance accurately from a model that ignores energy relaxation within the channel? The answer depends on the problem at hand. We can understand the relevant issues at least intuitively with a simple circuit model. Each energy channel can be modeled as a series of resistors, while inelastic flow is represented by the "vertical" resistors connecting nodes from different energies as sketched below.



If each energy channel has the same resistance profile then the vertical resistors carry no current since they connect nodes at the same potential. This is often the case in uniform conductors at low voltages, and so a model that ignores energy loss can be used to calculate resistance accurately, as long as momentum loss is modeled accurately for each energy cannel.

But in a conductor with a barrier in the middle the lower energies conduct poorly compared to the higher ones causing electrons to absorb energy from the lattice on the left of the barrier thus cooling the lattice leading to a mesoscopic Peltier effect. Such effects too are readily included within the NEGF method, but it requires us to include energy relaxing processes [8].



Structures with non-identical contacts like p-n junctions provide even more striking examples of the potential importance of vertical flow.

In short, inelastic processes may have a significant effect in specific problems especially at high bias. But it is reasonable at least to start a discussion of the physics of resistance without introducing inelastic interactions explicitly.

NEGF: OUR PERSPECTIVE

With this in mind, I would argue that in discussing steady-state current flow, it makes sense to start from coherent transport described by $\Sigma_{1,2}$, then introduce rigid interactions Σ_0 with different phase and momentum relaxing properties, and finally introduce Σ_0 for dissipative processes as needed. Even this last process can be described in the Born approximation using a one-electron picture, with subtle arguments related to the exclusion principle.

Note that this one-electron NEGF goes considerably beyond the scattering theory of transport which usually neglects interactions within the channel completely. Incoherence is commonly introduced in the scattering theory using an insightful observation due to Büttiker that dephasing processes essentially remove electrons from the channel and re-inject them just like real contacts and so one can include them phenomenologically by introducing conceptual contacts in the channel. This method is widely used in mesoscopic physics, but I am not aware of a convenient way to introduce pure phase relaxation without momentum relaxation if we wanted to.

In the NEGF method it is straightforward to choose Σ_0 so as to include arbitrary degrees of phase and momentum relaxation as we have seen. Indeed one could go further and introduce controlled amounts of other types of relaxation (spin for example) as well.

BUT SHOULD WE CALL THIS NEGF?

The answer to this question depends on which aspect of NEGF we consider fundamental:

A. Eqs.(3)-(5) or
B. Calculating the Σ's appearing in (3)-(5).

For historical reasons, these two aspects, A and B, are often intertwined in our thinking, but they need not be. Indeed these two aspects are completely distinct in the Boltzmann approach which is identified with the equation

$$\frac{\partial f}{\partial t} + \vec{v}.\vec{\nabla} f + \vec{F}.\vec{\nabla}_p f = \quad S_{op} f$$

and NOT with the evaluation of the scattering operator S_{op} which is analogous to the sigma's in (3)-(5). Boltzmann himself was unaware of the Fermi's golden rule commonly used nowadays to evaluate the S_{op} appearing in the equation bearing his name.

It thus seems appropriate to regard aspect A as more fundamental, and call Eqs.(3)-(5) the NEGF equations irrespective of how they are derived.

"CONTACT-ING" SCHRÖDINGER

In short, I feel that the scope and utility of Eqs.(3) - (4) transcends the MBPT-based approach originally used to derive it. It teaches us how to combine quantum dynamics with "contacts", much as Boltzmann taught us how to combine classical dynamics with "contacts", using the word "contacts" in a broad figurative sense to denote all kinds of irreversible processes.

The essence of NEGF is contained in Eqs.(3)-(5), while the actual evaluation of the Σ's may well evolve as we look at more and more different types of problems. The original MBPT–based approach may or may not be the most suitable, even for problems involving electron-electron interactions.

Above all we believe that by decoupling Eqs.(3)-(5) from the MBPT method originally used to derive them, we can make the NEGF method more transparent and accessible so that it can become a part of the standard training of physics and engineering students who need to apply it effectively to a wide

variety of basic and applied problems that require ***"connecting contacts to the Schrödinger equation" [9].***

Let me end by noting that the one-electron approach to NEGF is not just for students who lack the MBPT background. Even advanced students could benefit from it, because the one-electron approach provides an intuitive feeling whose value cannot be overstated especially as we venture into new frontiers like spintronics involving the control of electrons at a more delicate level. As Feynman remarked in his famous *Lectures on Physics*,

> " .. people .. say there is nothing which is not contained in the equations .. if I understand them mathematically inside out, I will understand the physics inside out. Only it doesn't work that way .. A physical understanding is .. absolutely necessary for a physicist."

REFERENCES

This is only a minimal list containing a few references that are explicitly mentioned in the text.

[1] (a) L.V.Keldysh (1965) Diagram Technique for non-equilibrium processes. *Sov.Phys. JETP* 20, 1018; (b) L.P.Kadanoff, G.Baym (1962) *Quantum Statistical Mechanics*, Frontiers in Physics Lecture Notes, Benjamin/Cummings.
[2] See for example, G.Stefanuci, R.van Leewuen, Non-Equilibrium Many-Body Theory of Quantum Systems, Cambridge (2013) and references therein.

[3] (a) S.Datta, Electron Transport in Mesoscopic Systems, Chapter 8, Cambridge (1995); (b) Quantum Transport, Atom to Transistor, Chapter 8-10, Cambridge (2005).
[4] S.Datta, *Lessons from Nanoelectronics,* World Scientific (2012). II Edition planned for 2016.

[5] S.Salahuddin, M.S.Lundstrom and S.Datta (2005) Transport Effects on Signal propagation in Quantum Wires, *IEEE Trans. Electron Dev*. **52,** 1734.
[6] See for example, P.Willke, T.Druga, R.G.Ulbrich, M.A.Schneider, M.Wenderoth (2015) Spatial Extent of a Landauer Resistivity Dipole in Graphene Quantified by Scanning Tunneling Potentiometry, *Nature Comm.***, 6,** 6399 and references therein.
[7] M.J.McLennan, Y.Lee and S.Datta (1991) Voltage Drop in Mesoscopic Systems, *Phys. Rev.* **B, 43**, 13846.
[8] R.Lake and S.Datta(1992) Energy Balance and Heat Exchange in Mesoscopic Systems, *Phys. Rev.* **B, 46**, 4757

[9] This is the approach used in my two-part online course *Fundamentals of Nanoelectronics*
https://nanohub.org/groups/u
based on [4], requiring only a familiarity with differential equations and linear algebra as prerequisites.