

Non-Equilibrium Green's Function Method in Matrix Representation and Model Transport Problems of Nanoelectronics

Yu.A. Kruglyak

Odesa State Environmental University, 15, Lvivska Str., 65016 Odesa, Ukraine

(Received 11 January 2013; published online 29 August 2013)

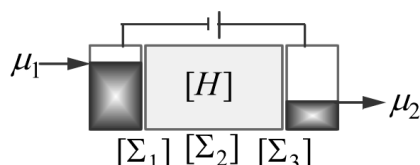
Non-equilibrium Green's functions method in matrix representation is presented and applied to model transport problems for 1D and 2D conductors using a nearest neighbor orthogonal tight-binding model in the frame of the «bottom-up» approach of modern nanoelectronics. Simple methods to account for electric contacts in Schrödinger equation to solve quantum electron transport problems are given.

Keywords: Nanophysics, Nanoelectronics, Molecular electronics, Bottom-up, Green's functions, NEGF method, Conductor modeling, 1D conductor, 2D conductor, Coherent transport, Quantum transport, Transmission coefficient.

PACS numbers: 73.23.Ad, 73.63. – b, 73.63.Rt

1. INTRODUCTION

Any nanoelectronics device has active conduction channel described by the Hamiltonian $[H]$. Conduction channel interacts with the source and drain, and with any of the contacts in a given specific device which stay in local equilibrium defined by the appropriate electrochemical potentials.



The interaction between the channel and the contacts are described by the self-energy contact matrices $[\Sigma_1]$ and $[\Sigma_2]$. Interaction of an electron in the channel with its environment is described by the self-energy matrix $[\Sigma_0]$, which in contrast to the matrices $[\Sigma_1]$ and $[\Sigma_2]$ is to be calculated self-consistently. The dimension of these square matrices is determined by the number N of the basis functions used for quantum-mechanical description of the conduction channel and contacts. Concrete form of the matrices is specified by the method used to solve the Schrodinger equation – semiempirical, based on the density functional theory or *ab initio*, and the choice of the basis functions. Once these matrices are composed the further procedure for calculating the conductivity, current and other electrophysical properties are straightforward which is the main purpose of this publication illustrated with the model transport problems of nanoelectronics having not only scientific but also pedagogical interest.

There are usually considered two limiting cases of electron transport – diffusional and ballistic. In the ballistic limit electron transport is controlled by the self-energy matrices $[\Sigma_1]$ and $[\Sigma_2]$, whereas the interactions inside of the channel are negligible. In contrast, in the diffusional limit the transport of electrons is controlled by interactions within the channel described by the matrix $[\Sigma_0]$, and the role of the contact matrices $[\Sigma_1]$ and $[\Sigma_2]$ is negligible. Not surprisingly, to about 1990

contacts were even not displayed on the charts. There is significant difference between Hamiltonian matrix $[H]$ and matrices $[\Sigma_{0,1,2}]$: Hamiltonian matrix represents conservative dynamic forces and is Hermitian, while the self-energy matrices accounts for entropic forces and are not Hermitian.

Classical description of transport processes is based on the Boltzmann transport equation. Non-Equilibrium Green's Function method (NEGF) is a quantum analogue of the Boltzmann equation; its foundations were laid by Martin and Schwinger [1], Kadanoff and Baym [2] and Keldysh [3]. Both approaches – the classical Boltzmann equation and quantum NEGF formalism are common in a sense that they both are taken into account the dynamic and entropic forces. In the ballistic limit, however, dynamic and entropic processes are spatially separated. Electrons skip from one contact to another one under the influence of only dynamic forces. Electrons inside the contacts are happen not be in equilibrium, but quickly come to equilibrium under the influence of entropic forces. This is the essence of the Landauer model for an elastic resistor proposed by Rolf Landauer in 1957 [4] long before its triumphal experimental confirmation on nanoresistors. Today it was indeed well established that ballistic resistors withstand fairly strong currents because Joule heating is negligible. Heat is released at the terminals, which due to their relatively massive quickly dissipate the heat. This separation of the dynamics from the thermodynamics to be one of the primary reasons that makes a bottom-up approach [5] starting with ballistic devices scientifically and pedagogically attractive.

2. EQUATIONS OF THE NON-EQUILIBRIUM GREEN'S FUNCTION METHOD

Our objective is to present the compact NEGF formalism with an account of the Landauer model for nanodevices. In our presentation we follow the concepts of Datta, Meir and Wingreen [5, 6] as the most appropriate for our purposes.

In our bottom-up approach we will start with elastic resistors for which energy exchange is confined to the contacts, and the problem of resistance can be treated

within a one-electron picture by connecting contacts to the Schrodinger equation

$$[H]\{\psi\} = E\{\psi\}$$

and add two more terms to it representing the outflow

$$[\Sigma] = [\Sigma_1] + [\Sigma_2]$$

and inflow from the contact

$$\{s\} = \{s_1\} + \{s_2\},$$

namely:

$$E\{\psi\} = [H]\{\psi\} + [\Sigma]\{\psi\} + \{s\},$$

where the Schrodinger equation is written directly in the matrix form, bearing in mind that the basis functions have been already chosen, so that the square matrices are shown in square brackets, and the column matrices – in curly brackets. Using this modified Schrodinger equation, the wave function can now be written in terms of the inverse matrix

$$\{\psi\} = [EI - H - \Sigma]^{-1}\{s\}$$

where I is unit matrix.

Matrix

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

is called Retarded Green's function and its Hermitian conjugate matrix

$$G^A = [G^R]^\dagger$$

is called Advanced Green's function. At the origin of these and other terms commonly used in the NEGF formalism we will not spend time now. We only note that the NEGF formalism applied to problems in nano-electronics is reduced to four equations, the first of which is the expression (1) for the Retarded Green's function.

Then the Schrodinger equation can be rewritten as

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

The product of the column $\{\psi\}$ by Hermitian conjugated row $\{\psi\}^\dagger$ gives

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

Non-equilibrium Green's function is defined as

$$G^n = 2\pi\{\psi\}\{\psi\}^\dagger,$$

so that the number of electrons is given by

$$N = \text{Tr} [G^n] / 2\pi.$$

Similarly inflow of electrons is described by

$$\Sigma^{in} = 2\pi\{s\}\{s\}^\dagger,$$

and now the non-equilibrium Green's function is

$$G^n = G^R \Sigma^{in} G^A \quad (2)$$

and serves as the second equation in the NEGF formalism.

The third equation is a matrix form of the density of

states $D(E)$, multiplied by 2π , and is called the spectral function A

$$2\pi D(E) = A = G^R I G^A = G^A I G^R = i[G^R - G^A], \quad (3)$$

where matrix $[I]$ is the anti-Hermitian part of the corresponding contact matrix

$$I = i[\Sigma - \Sigma^\dagger]$$

and describes the interaction of electrons in the channel with contacts.

The fourth equation of the NEGF formalism is the equation for the current through the terminal with number m

$$\tilde{I}_m = \frac{q}{h} \text{Tr} \left[\sum_m^{in} A - \Gamma_m G^n \right], \quad (4a)$$

which includes only those components of the matrices that are relevant to this terminal m . This is a specific current (per energy unit), it must still be integrated over the full range of energies to get the total current through the terminal m .

Lets transform equation (4a) as follows. Take into account (2) and (3), as well as

$$\Gamma = \sum_n \Gamma_n, \quad \Sigma^{in} = \sum_n \Sigma_n^{in}, \quad \Sigma_n^{in} = \Gamma_n f_n(E),$$

where $f_n(E)$ – the Fermi function of contact n . Then

$$\tilde{I}_m = \frac{q}{h} \sum_n \bar{T}_{mn} (f_m(E) - f_n(E)), \quad (4b)$$

where the transmission coefficient (transparency) between contacts m and n

$$\bar{T}_{mn} \equiv \text{Tr} \left[\Gamma_m G^R \Gamma_n G^A \right].$$

It is easy to prove a useful property of the transmission coefficient

$$\sum_n \bar{T}_{mn} = \sum_n \bar{T}_{nm} = \text{Tr} \left[\Gamma_m A \right].$$

So far only physical contacts $[\Sigma_{1,2}]$ in the quantum model of coherent transport have been considered, in which the electrons move coherently from source to drain through the channel described by the static Hamiltonian $[H]$ in the absence of interaction of the electrons with the environment $[\Sigma_0]$ along it moves through the channel. To account for the interaction $[\Sigma_0]$ from the formal point of view does not constitute any problem. All equations of the NEGF method (1)-(4) remain the same, but additional terms will appear in the matrices Σ , Γ , and Σ^{in}

$$\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_0,$$

$$\Gamma = \Gamma_1 + \Gamma_2 + \Gamma_0,$$

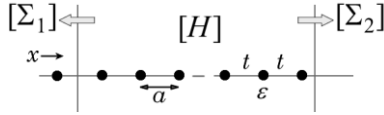
$$\left[\Sigma^{in} \right] = \left[\Gamma_1 \right] f_1(E) + \left[\Gamma_2 \right] f_2(E) + \left[\Sigma_0^{in} \right].$$

For any transport problem just to write the Hamiltonian $[H]$ and the self-energy matrices $[\Sigma]$. Once this is done, further calculations are performed by the NEGF method in standard way.

3. MODEL TRANSPORT PROBLEMS

3.1 1D Conductor

Consider the one-dimensional model of an infinite homogeneous conductor in the tight-binding approximation with the interaction t between only neighboring atoms in an orthogonal basis.



This approximation is known in quantum chemistry since 1931 as the Huckel molecular orbital method [7, 8]. Model parameters are Coulomb ε and resonance t integrals as well as the lattice parameter a . Even such a simple model correctly describes many properties of conjugated polyenes – $(CH = CH)_n$ $CH = [9-11]$, graphene [12], polyacetylenes and cumulenes $(C \equiv)_n C = [13]$, in the last, though, each carbon atom supplies two mutually orthogonal electrons π_y and π_z , which requires only an insignificant modification of the model.

Homogeneous 1D conductor is described by the standard Kronig – Penney theory of one-dimensional crystals and obeys the parabolic dispersion relation with the effective mass

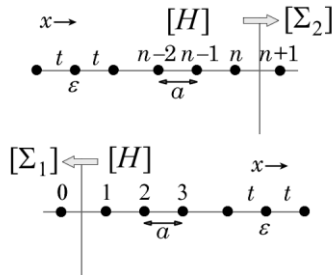
$$E = E_c + \frac{\hbar^2 k^2}{2m}.$$

In the low k values for the resonance and Coulomb integrals we have

$$E_c = \varepsilon + 2t, \quad -t \equiv t_0 \equiv \frac{\hbar^2}{2ma^2}.$$

It is straightforward now to write down the $[H]$ matrix with ε on the diagonal and t on the upper and lower diagonals. What needs discussion are the self-energy contact matrices. The basic idea is to replace an infinite conductor described by the Hamiltonian $[H]$ with a finite conductor described by $[H + \Sigma_1 + \Sigma_2]$ assuming open boundary conditions at the ends, which means that electron waves escaping from the end surface do not give rise to any reflected waves, as a good contact should ensure.

For a one-dimensional lattice the idea is easy to see.



Let the conductor has a limited length of n atoms, numbered from 1 to n . Left contact 1 starts before the atom chain with the number 1, and the right contact 2 – after the atomic chain with the number n . Contacts

have no incoming streams, only outgoing ones.

In the n -th row

$$E\Psi_n = t\Psi_{n-1} + t\Psi_{n+1}$$

of the Schrodinger equation

$$E\Psi_n = \sum_m H_{nm} \Psi_m$$

term $t\Psi_{n+1}$ already belongs to terminal 2, which according to the equation

$$E = t \frac{\Psi_{n-1}}{\Psi_n} + \varepsilon + t \frac{\Psi_{n+1}}{\Psi_n},$$

contributes to the energy equal to $t\Psi_{n+1} / \Psi_n$. This energy is the self-energy of contact 2. We have

$$\Psi_{n+1} = \Psi_n e^{ika}$$

thus

$$E\Psi_n = t\Psi_{n-1} + (\varepsilon + te^{ika})\Psi_n,$$

in other words the effect of the contact is simply to add $t \exp(+ika)$ to H_{nn} which amounts to adding the self-energy

$$\Sigma_2 = \begin{bmatrix} \ddots & & \dots & \\ & 0 & 0 & 0 \\ \dots & 0 & 0 & 0 \\ & 0 & 0 & te^{ika} \end{bmatrix}$$

to the Hamiltonian. Note the only non-zero element is the (n, n) element.

The same self-energy has the contact 1, and its value in contact matrix is placed as element $(1, 1)$

$$\Sigma_1 = \begin{bmatrix} te^{ika} & 0 & 0 & \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \\ & \dots & & \ddots \end{bmatrix}.$$

Remaining elements of matrices Σ_1 and Σ_2 are zero. In short, the self-energy function for each contact has a single non-zero element corresponding to the point that is connected to that contact.

Energy matrices H , Σ_1 , and Σ_2 are written down, next we calculate the Retarded Green's function G^R , Advanced function G^A , matrices Γ_1 и Γ_2 and, finally, transmission coefficient T_{12} and conductivity $G(E)$

$$G(E) = \frac{q^2}{h} Tr [\Gamma_1 G^R \Gamma_2 G^A] = \frac{q^2}{h} \bar{T}_{12}.$$

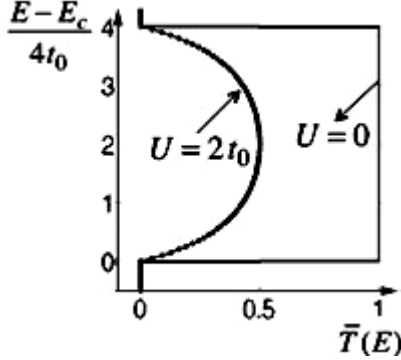
A good test case for any theory of coherent quantum transport is the conductance function for a length of uniform ballistic conductor. If we are doing things right, the conductance function $G(E)$ should equal the quantum of conductance q^2 / h times an integer equal to the number of modes $M(E)$ which is one for 1D conductors neglecting spin. This means that the transmission should equal one over the energy

$$0 < E - E_c < 4t_0,$$

covered by the dispersion relation

$$E = \varepsilon + 2t \cos ka = E_c + 2t_0(1 - \cos ka),$$

but zero outside this range as shown below with $U = 0$.



Another good example is that of a conductor with just one scatter whose effect is included in the Hamiltonian $[H]$ by changing the diagonal element corresponding to that point to $\varepsilon + U$. Transmission through a single point scatter in a 1D wire with $U = 2t_0$ is also shown above.

Let's calculate the density of states $D(E)$ of 1D conductor. The number of states for 1D conductor of length L that have a momentum less than a given value p

$$N(p) = \frac{2L}{h/p},$$

so that the density of states

$$D(E) = \frac{dN}{dE} = \frac{2L}{h} \frac{dp}{dE} = \frac{L}{\pi \hbar v},$$

where it was taken into account that for the isotropic dispersion $E(p)$ speed $v = dE / dp$.

Let's get the same expression for the density of states using the NEGF method. For a homogeneous 1D conductor it is sufficient to consider only one atom in the chain, as it plays the role of the unit cell. In this case the lattice constant a plays the role of the conductor length L . For Retarded Green's function we have

$$G^R = [E - \varepsilon - 2te^{ika}]^{-1}.$$

Presenting exponent through the sine and cosine and considering the dispersion relation above we find

$$G^R = i / 2t \sin ka.$$

We also have

$$\hbar v = \frac{dE}{dk} = -2at \sin ka,$$

where the first equality follows from the isotropy of the 1D conductor, and the second follows from the dispersion relation. For G^R we finally have

$$G^R = \frac{i}{2t \sin ka} - \frac{-i}{\hbar v / a},$$

and Advanced Green's function

$$G^A = \frac{i a}{\hbar v}.$$

Spectral function

$$A = i[G^R - G^A] = \frac{2a}{\hbar v},$$

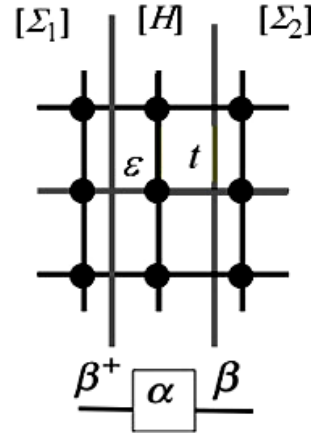
and density of states

$$D(E) = \frac{A}{2\pi} = \frac{a}{\pi \hbar v}$$

coincides with the already given above expression.

3.2 2D Conductor

Among the seminal experiments from the 1980's that gave birth to mesoscopic physics was the observation that the conductance of a ballistic 2D conductor went down in integer multiples of $2q^2 / h$ as the width of the narrow region was decreased. To understand this class of experiments we need a 2D model as simple as possible. Two-dimensional model of unlimited homogeneous conductor in the Huckel approximation is sufficient.



The model parameters ε and t are chosen in a way as to satisfy the standard dispersion relation with the effective mass. Finally

$$E_c = \varepsilon + 4t, \quad -t \equiv t_0 \equiv \frac{\hbar^2}{2ma^2}.$$

Constructing Huckel Hamiltonian H is not difficult, it is necessary to discuss only the structure of the contact matrices for 2D conductor. Let conductor have p atoms across the conductor width and q atoms along the conductor length, thus the conductor matrix has the form $(p \times q)$. Such a 2D conductor can be conventionally represented as p 1D conductors connected in parallel, each of length q . Matrix $(p \times q)$ is the row matrix of length q with elements as column matrices of

length p . The figure above shows a conductor with one column of the form $(p \times 1)$. Each of the q columns is described by its own Huckel matrix α of order p . For example, for $p = 3$ it looks as

$$\alpha = \begin{bmatrix} \varepsilon & t & 0 \\ t & \varepsilon & t \\ 0 & t & \varepsilon \end{bmatrix}.$$

Columns are connected to each other by resonance integrals t in the following way. Consider the connection between the columns with neighboring numbers n and $n + 1$. This connection is described by the scalar matrix $\beta = t \cdot I$ of order p , where I is unit matrix. In our model, naturally, $\beta = \beta^+$. For example, for $p = 3$

$$\beta = \begin{bmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t \end{bmatrix}.$$

Hamiltonian H has a block structure. The same matrices α of order p are located on its main diagonal, and next to it above and below the diagonals are filled with matrices β of the same order p , the rest of elements are zero. If the length of the conductor is, say, $q = 10$ atoms, and the width of the conductor is $p = 5$ atoms, the order of H is $p \times q = 50$.

The solution of the eigenvalue problem of the Hamiltonian H is reduced to diagonalization of $[a]$

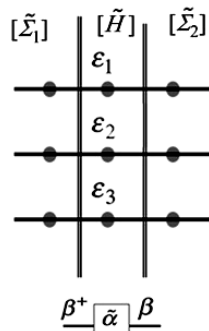
$$[\tilde{\alpha}] = [V]^+ [\alpha] [V],$$

where $[V]$ is a matrix whose columns represent the eigenvectors of matrix $[a]$, thus

$$\tilde{\alpha} = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{bmatrix}.$$

Matrix β is not affected by the basis transformation, because it is already diagonal. Diagonalisation of $(p \times p)$ matrix a leads to the vanishing of the resonance integrals t , connecting rows of the initial Hamiltonian matrix, that is, to transform 2D conductor to p 1D conductors connected in parallel each of q atoms in length with energies $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_p$ equal to eigenvalues of matrix a

$$\varepsilon_n = \varepsilon - 2t_0 \cos k_n a, \quad k_n a = \frac{n\pi}{p+1}.$$



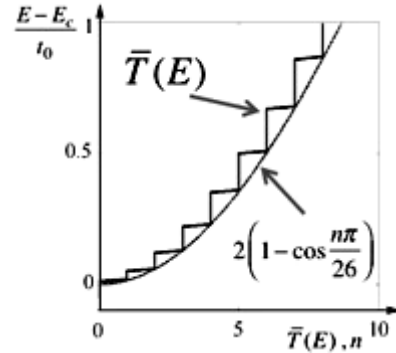
Each of these p parallel 1D conductors has a transmission of one in the energy range $(t_0 \equiv |t|)$

$$\varepsilon_n - 2t_0 < E < \varepsilon_n + 2t_0.$$

Adding all the individual transmissions for all p modes of a conductor we obtain the transmission showing up-steps in the lower part (see next figure) and down-steps in the upper part (not shown). Usually when modeling n-type conductors we use the lower part of the band as shown below, and so we see only the up-steps occurring at $\varepsilon_n - 2t_0$. The ε_n are the eigenvalues of α which are given by

$$\varepsilon_n - 2t_0 = E_c + 2t_0 \left(1 - \cos \frac{n\pi}{p+1} \right).$$

The results of calculation of the transmission coefficient by the NEGF method when the number of atoms in conductor width $p = 25$ are shown together with the step round calculated from the previous formula with the same width of the conductor $p = 25$.



The approach we just described of viewing a 2D/3D conductor as a set of 1D conductors in parallel looks to us not only physically correct but also as a very powerful tool for interpretation of experimental data. Each of these 1D conductors is called a mode (or subband) and has a dispersion relation

$$E_n(k_x) = \varepsilon_n - 2t_0 \cos k_x a.$$

Let us now address the question how do we write the self-energy matrices for the contacts. Ideally the contact regions allow electrons to exit without any reflection from the contact borders and with this in mind a simple way to evaluate $[\Sigma]$ is to assume the contacts to be just uniform extensions of the channel region.

The viewpoint we just discussed allows us to picture a 2D conductor as a set of decoupled 1D conductors by converting from the usual lattice basis to an abstract mode basis through a basis transformation

$$[\tilde{X}] = [V]^+ [X] [V]$$

with X being any matrix in the regular lattice basis. A unitary transformation like this can be reversed by transforming back

$$[X] = [V] [\tilde{X}] [V]^+$$

as it was demonstrated above for Hamiltonian H .

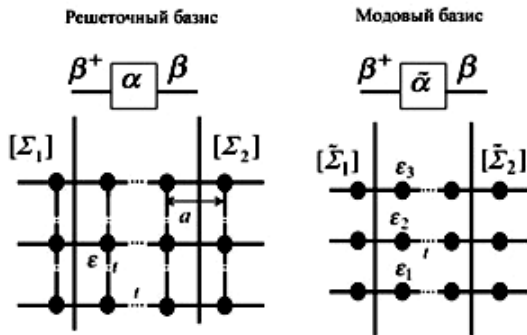
As a result, for each of the p independent 1D conductors we can easily write down the self-energy in the mode basis and then connect them together to the total Σ matrix.

In the Huckel model of 2D conductor each of p 1D conductors is characterized by contact self-energy $t \exp(ika)$, with the appropriate ka for that 1D conductor at a given energy E . For mode n we have

$$E = \varepsilon_n - 2t_0 \cos k_n a,$$

so that overall we could write the contact matrix in the mode basis as

$$[\tilde{\Sigma}_1] = \begin{bmatrix} te^{ik_1 a} & & \dots \\ & te^{ik_2 a} & \\ \dots & & \ddots \\ & & & te^{ik_n a} \\ & & & & \ddots \end{bmatrix},$$



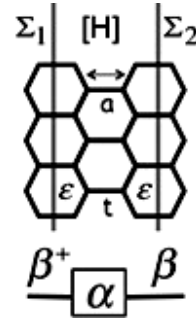
and then transform it back to the lattice basis

$$[\Sigma_1] = [V][\tilde{\Sigma}_1][V]^+$$

The method of basis transformation given above is based on a physical picture that is very powerful and appealing. However, it can not always be used at least not as straightforwardly since in general it may not be possible to diagonalize both α and β matrices simultaneously.

Finally, we formulate a general method for constructing self-energy contact matrices. Any 2D conductor with a uniform cross-section along the entire length of the conductor can be broken into fragments each having an on-site matrix Hamiltonian $[a]$ that mimic each other along the length of the conductor and cou-

pled to the next fragment by a matrix $[\beta]$ as shown below for non-uniform graphene ribbon with its two-atomic rhombic unit cell and 12 atoms fragment. Each



of these matrices is of size $(n \times n)$, n being the number of basis functions describing each fragment.

Let's look at the right border of the conductor with the contact. The self-energy matrix is zero except for the last $(n \times n)$ block at the contact surface

$$\Sigma_2(E) = \begin{bmatrix} \dots & \dots & \dots \\ \dots & 0 & 0 \\ \dots & 0 & \beta g_2 \beta^+ \end{bmatrix}$$

The non-zero block is given by $\beta g_2 \beta^+$ where g_2 is called the surface Green's function for contact 2, and is obtained by iteratively solving the equation for g_2

$$[g_2]^{-1} = (E + iO^+)I - \alpha - \beta^+ g_2 \beta,$$

where O^+ represents a positive infinitesimal being used to control the convergence of iteration process and iO^+I ensures that a numerical iterative solution converges on the solution for which Γ has all positive eigenvalues.

Validity and deviation of the four basic equations of the NEGF method in matrix representation in the bottom-up approach as well as detailed discussion of single- and multilevel resistors in semiclassical and quantum models, 1D conductor with one scatterer, graphene, and deviation of the general method given above to obtain the self-energy contact matrices is supposed to be submitted for publication to the Journal of Nano- and Electronic Physics.

REFERENCES

1. P.C. Martin, J. Schwinger, *Phys. Rev.* **115**, 1342 (1959).
2. L.P. Kadanoff, G. Baym, *Quantum Statistical Mechanics* (New York: W.A.Benjamin: 1962).
3. L.V. Keldysh, *JETP* **47**, 1515 (1964).
4. Rolf Landauer, *IBM J.Res.Dev.* **1**, 223 (1957).
5. Supriyo Datta, *Lessons from Nanoelectronics: A New Perspective on Transport* (Hackensack, New Jersey: World Scientific Publishing Company: 2012).
6. Y. Meir, N.S. Wingreen, *Phys. Rev. Lett.* **68**, 2512 (1992).
7. Yu.A. Kruglyak, et al., *Computational Methods in Quantum Chemistry. Calculation of π -electronic structure of molecules by simple molecular orbitals methods* (Kiev: Naukova Dumka: 1967) [In Russian].
8. Yu.A. Kruglyak, et al., *Computational Methods of Molecular Structure and Spectra* (Kiev: Naukova Dumka: 1969) [In Russian].
9. Yu.A. Kruglyak, I.I. Ukrainsky, *Int. J. Quantum Chem.* **4**, 57 (1970).
10. I.I. Ukrainsky, Yu.A. Kruglyak, *Ukr. Fiz. Zh.* **15**, 1068 (1970).
11. G.F. Kventsels, Yu.A. Kruglyak, *Theor. Chim. Acta* **12**, 1 (1968).
12. Yu.A. Kruglyak, N.E. Kruglyak, *Visnyk Odessa State Environmental Univ.* No 13, 207 (2012).
13. Yu.A. Kruglyak, G.G. Dyadyusha. *Theor. Chim. Acta* **10**, 23; **12**, 18 (1968).