On the implementation of a position-dependent charge carrier mass in NEGF simulations

Fabio Leão
Department of Electronics and Systems
Universidade Federal de Pernambuco
Recife, Brazil
fabio.leao@ufpe.br

Diego Neves
Department of Communication
Universidade Estadual de Campinas
Campinas, Brazil
ORCID 0000-0002-6092-0592

Stefan Blawid
Centro de Informática
Universidade Federal de Pernambuco
Recife, Brazil
ORCID 0000-0001-6982-4575

Abstract—Semiconductor devices are integrated from a limited number of fundamental building blocks. Important structures like ohmic contacts, band-to-band tunneling barriers or quantum wells are derived from metal-semiconductor interfaces, doping transitions or heterojunctions. Current injection across these material intersections is reduced due to a change in the effective mass of the charge carriers. Thus, mass discontinuities need to be carefully modeled in device simulations. In this tutorial brief, we give an overview of possible finite difference schemes gathering information that might be difficult to encounter in the original literature. Different discretized device Hamiltonians are used to compute the transmission coefficient with the help of non-equilibrium Green's functions. Two approaches are validated by comparing numerical and analytical transmission values at mass discontinuities in planar bands and potential barriers. We show that a smooth material transition leads to an improved transmission coefficient.

Index Terms—Heterojunction, NEGF, Effective mass approximation, Finite difference method, Transmission coefficient

I. INTRODUCTION

Technology computer-aided design (TCAD) is an essential tool when developing semiconductor manufacturing platforms for integrated electronic circuits. In particular, the virtual design of electron devices on a computer allows both, architecture optimization for manufacturing (Design Technology Co-Optimization - DTCO) and the derivation of compact models for circuit simulation (Electronic Design Automation - EDA). Device TCAD is based on fundamental transport equations of charge carriers in semiconductors. Semiconductor models are hierarchical and range from macroscopic semi-classical to microscopic quantum equations. For open quantum systems, like nanoscale electron devices, the density matrix formulation of quantum mechanics has certain advantages since it allows an easier implementation of transparent boundary conditions and of scattering between charge carriers and phonons.

An approach to compute the device density matrix are non-equilibrium Green's functions (NEGFs) [1]. In essence, NEGFs treat the contact eigenstates as incoherent device excitations and transform each into a corresponding device wave function. Recent applications of NEGFs include the optimization of heterojunction tunnel field-effect transistors (FETs)

with tunneling along the gate field [2]–[4], electrostatically doped FETs that can be reconfigured to operate in multiple modes [5], [6], photodiodes using a potential barrier to block the flow of majority carriers [7], and quantum cascade lasers [8]. In all these devices charge carriers are either contained-in or controllably-released-from engineered quantum wells. In general, band engineering requires the use of different materials forming heterojunctions like $MoS_2/MoTe_2$, $Ge/SiGe$, $GaAs/AlGaAs$, $InAs/GaSb$ etc.

We revisit the modeling of charge transport across a heterojunction focusing on the mass discontinuity of the carriers. The wave vector mismatch at the mass discontinuity leads to reflections of the carrier waves and the impact can be significant [9]. Thus, it is well recognized that the mass mismatch needs to be taken into account in numerical simulations. We report here on the convergence behavior of two different discretization schemes for the Hamiltonian with a mass discontinuity. The results reported here are part of the development of a Python NEGF module by the authors, which will be distributed publicly (details reported elsewhere).

II. ESSENTIAL EQUATIONS

In a ballistic device, the solution $\psi(x)$ to the one-particle Schrödinger equation $\hat{H}\psi = E\psi$ describes electrons flowing through a conducting channel. When the carrier effective mass is position-dependent, the Hamiltonian operator is given by

$$\hat{H}\psi(x) = -\frac{\hbar^2}{2}\frac{d}{dx}\left(\frac{1}{m(x)}\frac{d\psi}{dx}\right) + U(x)\psi(x), \quad (1)$$

where $m(x)$ and $U(x)$ are the effective mass and the potential energy at x. The first step to compute an approximation of ψ or to use the NEGF method is to employ a discretization scheme. In this case, a square matrix H represents the operator $\hat{H}$ within the channel. The explicit form of H is discussed in the following section.

One of the main advantages of the NEGF method is that influences of the contacts are modelled by self-energy matrices, Σ_s and Σ_d, that have the same size as H. Here sub-indexes s and d indicate quantities related to left- and right-sided contacts, respectively. The retarded matrix Green function is computed as

$$G(E) = [(E + j\eta)I - H - \Sigma_s(E) - \Sigma_d(E)]^{-1}, \quad (2)$$

where I is the identity matrix, j is the imaginary unit, and η is an extremely small positive number. All the quantities of interest can be derived from G. In particular, we can obtain the transmission coefficient

$$T(E) = \mathrm{Tr}(\Gamma_s(E)A_d(E)), \tag{3}$$

and the local density of states

$$\mathrm{LDOS}(x_i, E) = \frac{A_{s,ii}(E) + A_{d,ii}(E)}{2\pi}, \tag{4}$$

where $\Gamma_{s/d} = -j\left[\Sigma_{s/d} - \Sigma_{s/d}^\dagger\right]$ and $A_{s/d} = G^\dagger\Gamma_{s/d}G$. The explicit form of $\Sigma_{s/d}$ used in this work is the same as in [5]. From the expressions above, the device density matrix, the charge density and the electric current can be computed.

III. DISCRETIZATION OF THE HAMILTONIAN

In a discrete version of the Schrödinger equation, the vector $\psi = [\psi(x_1), \psi(x_2), \ldots, \psi(x_N)]^T$ represents the wave function, where $\{x_1, x_2, \ldots, x_N\}$ is a set of N positions. We consider a uniform discretization in which $a = x_{i+1} - x_i$ is constant. In addition to this, the Hamiltonian operator in Eq. (1) translates into a square matrix H such that $H\psi \approx [\hat{H}\psi(x_1), \hat{H}\psi(x_2), \ldots, \hat{H}\psi(x_N)]^T$. To obtain H, some kind of finite difference formula must be applied to the kinetic energy term $\frac{d}{dx}(\psi'(x)/m(x))$. Applying such formulas to the expansion

$$-\frac{m'(x)}{m^2(x)}\psi'(x) + \frac{1}{m(x)}\psi''(x)$$

of this term has been said to lead to convergence issues [10]. The behavior of $m'(x)$ at a mass step is possibly the main reason. A more reliable approach is to first discretize $\frac{d}{dx}\phi(x)$, where $\phi(x) = \psi'(x)/m(x)$, and then discretize the resulting terms. Throughout this text, we use and compare two different discretization schemes obtained in this way and denoted by S_1 and S_2.

Both schemes give rise to a tridiagonal Hamiltonian matrix, whose diagonal and off-diagonal elements are given by

$$H_{ii} = t_i^- + t_i^+ + U_i \text{ and} \tag{5}$$
$$H_{i,i\pm1} = -t_i^\pm, \tag{6}$$

respectively, with $i = 1, 2, \ldots, N$. The hopping parameters are defined as

$$t_i^\pm = \frac{\hbar^2}{2a^2 m_i^\pm}, \tag{7}$$

where the mass average $m_i^\pm$ is specific of each scheme and depends on $m(x_i)$ and $m(x_{i\pm1})$. For S_1, the mass average is given by

$$m_i^\pm = \frac{m_i + m_{i\pm1}}{2}. \tag{8}$$

This is the discretization scheme commonly employed to solve the Schrödinger equation [11], [12] or to compute NEGFs [7], [13]. In this brief, we introduce a new discretization scheme, S_2, in which

$$m_i^\pm = \left[\frac{1}{2}\left(\frac{1}{m_i} + \frac{1}{m_{i\pm1}}\right)\right]^{-1}. \tag{9}$$

Besides being deduced in a similar way, these schemes lead to very different values of the elements in Eqs. (5) and (6) when $m_i/m_{i+1} \gg 1$ or $m_i/m_{i+1} \ll 1$. Therefore, it is interesting to compare how this pair of schemes behave when there are large mass steps in the channel.

The scheme S_1 related to Eq. (8) results from using a central finite-difference formula twice followed by a variable change [10]. In the first step, we find

$$\left.\frac{d\phi}{dx}\right|_{x_i} \approx \frac{\phi(x_i + a) - \phi(x_i - a)}{2a},$$

where each term is given by

$$\phi(x_i \pm a) = \pm\frac{1}{m(x_i \pm a)}\left(\frac{\psi(x_i \pm 2a) - \psi(x_i)}{2a}\right).$$

Making the transformation $2a \to a$ and rearranging the terms, we get

$$\left.\frac{d\phi}{dx}\right|_{x_i} \approx \frac{1}{(2a)^2}\left(\frac{\psi(x_i + a) - \psi(x_i)}{m(x_i + a/2)} - \frac{\psi(x_i) - \psi(x_i - a)}{m(x_i - a/2)}\right).$$

The mass $m_i^\pm$ in Eq. (8) approximates $m(x_i \pm a/2)$ by means of linear interpolation.

On the other hand, the scheme S_2 related to Eq. (9) results from the average of forward and backward finite-difference formulas:

$$\left.\frac{d\phi}{dx}\right|_{x_i} \approx \frac{\phi(x_i + a) - \phi(x_i)}{a}$$

and

$$\left.\frac{d\phi}{dx}\right|_{x_i} \approx \frac{\phi(x_i) - \phi(x_i - a)}{a}.$$

Each of the above expressions is again discretized using such formulas. In this step, the choice between forward and backward rules is made to ensure that H is a hermitian matrix. As a result, we get the average of inverses as the mass $m_i^\pm$ in Eq. (9).

Another possible discretization scheme is described in [14] (See Eq. (7.1.5) and the following discussion). The proposed scheme produces a Hamiltonian matrix analogous to that of S_2. However, it is restricted to abrupt mass steps. On top of that, a position point must be located at the interface where the mass changes.

IV. ANALYTIC CASE STUDY

Eq. (1) implies that both ψ and $1/m \times \mathrm{d}\psi/\mathrm{d}x$ need to remain continuous at a mass discontinuity. With this boundary condition the transmission coefficient for a piecewise constant conduction band energy E_i can be calculated analytically. The index i indicates the different "pieces" of the band profile. A quantum well leads to transcendent equations for the electron wave numbers:

$$k_i = \sqrt{2m_i(E - E_i)/\hbar^2} \tag{10}$$

that need to be solved to find the "discrete" well states, see, e.g., section 2.7.2 in [10]. Such simplified band profiles are very useful to validate numerical algorithms, which then can be applied to more realistic heterojunctions.

Here, we use a wall (or sink) in the conduction band with mass discontinuities at the edges instead of a well. The transmission $T(E)$ through a double barrier (a well) can be deduced from the transmission of a single one. A potential wall/sink can also serve as a zero-order transistor model. The $T(E)$ through a potential barrier can be solved in closed form and is described in [15]. The conduction band E_i consists of three "pieces": two access regions denoted source, $i = s$, and drain, $i = d$, and a connecting wire of length L_{ch} denoted channel, $i = ch$. The resulting transmission coefficient reads:

$$T = \frac{b}{\frac{1}{4}(1+b)^2 + \frac{1}{4a^2}[b^2 - (1+b^2)a^2 + a^4]\sin^2(k_{ch}L_{ch})} \quad (11)$$

with

$$M = \frac{m_s}{m_{ch}}, \; a = M\frac{k_{ch}}{k_s}, \; b = \frac{k_d}{k_s} \; . \quad (12)$$

To validate different implementation schemes for mass discontinuities, we should choose (unrealistically) large values for the mass ratio M and look for quantities that are especially sensitive. Such a quantity is the minimum transition coefficient:

$$T_{\min}(E) = \frac{4a^2b}{(a^2+b)^2} \; . \quad (13)$$

The difference between the analytic and numeric results for $T_{\min}$ will be used to benchmark the two different implementation schemes described in Sec. III. More precisely, a single valued parameter $T(E_{\min})$ with

$$E_{\min} \approx E_{ch} + \frac{9\pi^2}{4L_{ch}^2}\frac{\hbar^2}{2m_{ch}} \quad (14)$$

(the approximative energy of the first transmission minimum, ca. 7% error) will be computed. In the benchmark example, the values in Tab. I are used.

TABLE I
POTENTIAL BARRIER PARAMETERS FOR BENCHMARKING DIFFERENT DISCRETIZATION SCHEMES OF THE HAMILTONIAN.

$m_{s/d}$	m_{ch}	$E_{s/d}$	E_{ch}	L_{ch}	$E_{\min} - E_{ch}$
Mm_{ch}	$0.05m_e$	-0.4 eV	+0.3 eV	30 nm	19 meV

V. RESULTS

For a given step-like potential and effective mass profile along the device channel, Eq. (1) was solved using Pyhton's numpy module. From $G(E)$ the transmission coefficient and the local DOS are obtained. The developed solution code runs on a common computer notebook in a few minutes. Details of the solution code will be reported elsewhere.

Fig. 1 displays the LDOS and transmission coefficient of a device with a barrier of height $\Delta E = 0.7$ eV and $M = 2$. The LDOS is measured in number of states per energy and the color scale is adjusted to the LDOS averaged over the device. The result shows that the differences in electrostatic potential and effective masses between the source/drain and the channel regions generate harmonic states in the channel, similar to
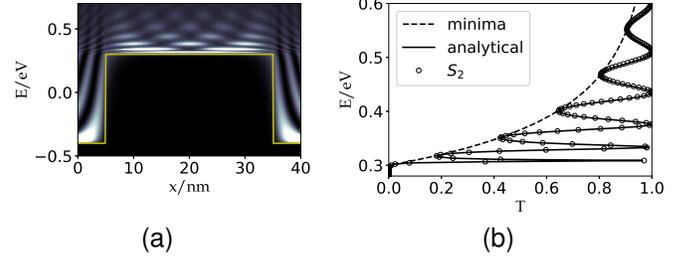
Fig. 1. (a) Local Density of States for $M = 2$. (b) Comparison between the analytical solution (line) with its minimal curve (dashed) and the numeric solution (open circles) obtained for the discretization scheme S_2.
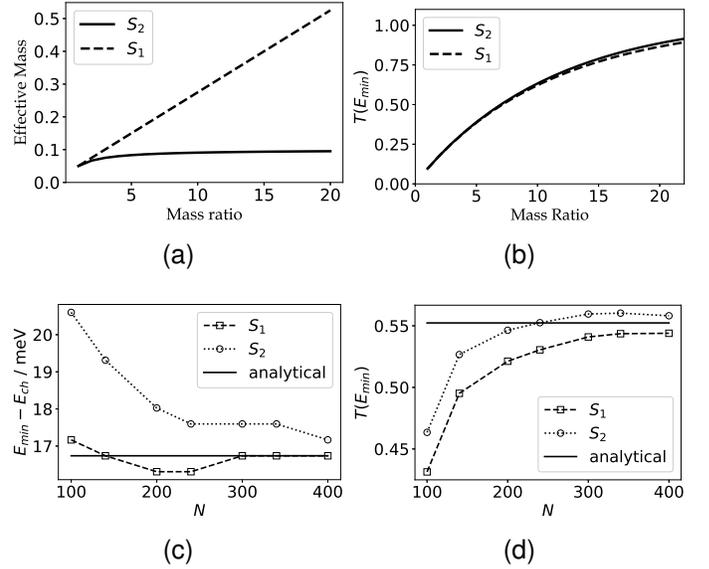
Fig. 2. Comparison of the two discretization schemes S_1 and S_2: (a) Effective mass at the transition point with increasing M computed using Eq. (8) and Eq. (9) for S1 and S2, respectively, (b) M-dependence of the transmission coefficient $T(E_{\min})$ at the energy of the first minima obtained from the analytic curve, $E_{min} - E_{ch} \approx 16.7$ meV, (c) Numeric convergence of $E_{\min}$ with the number of spatial points N at $M = 8$, (d) Numeric convergence of $T(E_{\min})$ with the number of spatial points N at $M = 8$.

a potential well. The transmission of charge carriers with energies below 0.3 eV is strongly suppressed by the $30\,\text{nm}$ barrier. Electrons with energies above 0.3 eV are crossing the channel. The transmission coefficient oscillates between 1 and $T_{\min}$. The numerical solution fits the analytical one validating the here newly proposed discretization scheme S_2.

Results for the two discretization schemes S_1 and S_2 are compared in Fig. 2. The effective masses in both schemes differ significantly. The S_1 scheme increases the effective mass linearly with M, while in S_2 the effective mass approaches $2m_{ch}$. Regardless these differences in $m_i^\pm$, both schemes predict the first transmission minima correctly. However, the numeric convergence differ considerably. The convergence of $E_{\min}$ and T with higher spatial resolution is faster in S_1 and S_2, respectively.

The transmission of charge carriers in a planar band can be

matter for high-resolution spatial sampling, significant differences are obtained for finite step sizes. The transmission coefficient might be over- or underestimated depending on the charge carrier energy. This is an important observation since high-resolution sampling comes at high computational cost. In this tutorial brief we gave an overview on how to derive and validate possible finite difference equations. The work is part of a continuous effort to implement an educational module for the Python programming language that enables the simulation of electron devices for given band profiles. As an application example, we showed that smoothing an abrupt mass discontinuity leads to largely improved transmission coefficients.

Fig. 3. Charge carrier transmission in a planar band with mass discontinuity: LDOS (a) and transmission (b) with no mass change $M = 1$, LDOS (c) and transmission (d) with abrupt mass change $M = 20$, LDOS (e) and transmission (f) for a smooth change of mass to $M = 20$ throughout the device with constant increase of mass over different intervals ranging from $0.2\,\mathrm{nm}$ to $20\,\mathrm{nm}$. The mass increases linearly in the x region $20\,\mathrm{nm} \pm \mathrm{interval}/2$ from $m = 0.05 m_\mathrm{e}$ to m_e. The mass profile along the device channel is indicated in (e) as dashed line. The color scale for the LDOS has been changed in (e) to highlight the energy and position dependence. Values above $1\,\mathrm{eV}^{-1}$ are depicted as white.

significantly reduced by mass discontinuities. Fig. 3 displays the LDOS and $T(E)$ for the ideal case (a,b), for an abrupt mass change in the middle of a channel (c,d) and for a smooth mass change within a finite transition region (e,f). The transmission coefficient for an abrupt mass change can be computed analytically $T_\mathrm{max} = 4\sqrt{M}/(1 + \sqrt{M})^2$ and is reduced from $T(E, M = 1) = 1$ to $T(E, M = 20) = 0.6$. A smooth transition to $M = 20$ over an $20\,\mathrm{nm}$ interval increases the transmission again to nearly one, i.e., $T(E = 0.1\,\mathrm{meV}) = 0.97$.

VI. Conclusion

At a heterojunction both the charge carrier wave function and its derivative normalized to the effective mass need to remain continuous. The device Hamiltonian has to be discretized correspondingly. The resulting finite difference schemes exhibit very different convergence behaviors. Although the choice of the finite difference method does not

References

[1] S. Datta, "Nanoscale device modeling: the Green's function method," *Superlattices and Microstructures*, vol. 28, no. 4, pp. 253 – 278, 10 2000.

[2] C. Grillet, A. Cresti, and M. G. Pala, "Vertical GaSb/AlSb/InAs Heterojunction Tunnel-FETs: A Full Quantum Study," *IEEE Transactions on Electron Devices*, vol. 65, no. 7, pp. 3038–3044, 2018.

[3] A. Afzalian, G. Doornbos, T.-M. Shen, M. Passlack, and J. Wu, "A High-Performance InAs/GaSb Core-Shell Nanowire Line-Tunneling TFET: An Atomistic Mode-Space NEGF Study," *IEEE Journal of the Electron Devices Society*, vol. 7, pp. 88–99, 2018.

[4] Y. Balaji, Q. Smets, A. Szabo, M. Mascaro, D. Lin, I. Asselberghs, I. Radu, M. Luisier, and G. Groeseneken, "MoS2/MoTe2 Heterostructure Tunnel FETs Using Gated Schottky Contacts," *Advanced Functional Materials*, vol. 30, no. 4, p. 1905970, 2020.

[5] S. Blawid, D. Neves, and R. Moura, "Exploring Quantum Tunneling in Ultrathin Transistors with Multiple Top Gates," in *2020 IEEE Latin America Electron Devices Conference (LAEDC)*, vol. 00, 2020, pp. 1–4.

[6] R. Moura, M. Claus, and S. Blawid, "Multi-mode nanoFETs: From low power to high performance on demand," *Semiconductor Science and Technology*, vol. 35, no. 5, p. 055027, 04 2020.

[7] N. D. Akhavan, G. A. Umana-Membreno, R. Gu, J. Antoszewski, and L. Faraone, "Optimization of Superlattice Barrier HgCdTe nBn Infrared Photodetectors Based on an NEGF Approach," *IEEE Transactions on Electron Devices*, vol. 65, no. 2, pp. 591–598, 2018.

[8] T. Grange, D. Stark, G. Scalari, J. Faist, L. Persichetti, L. D. Gaspare, M. D. Seta, M. Ortolani, D. J. Paul, G. Capellini, S. Birner, and M. Virgilio, "Room temperature operation of n-type Ge/SiGe terahertz quantum cascade lasers predicted by non-equilibrium Green's functions," *Applied Physics Letters*, vol. 114, no. 11, p. 111102, 2019.

[9] J. R. Villa-Angulo, R. Villa-Angulo, K. Solorio-Ferrales, S. E. Ahumada-Valdez, and C. Villa-Angulo, "Effect of effective mass mismatch in CdS/CdTe heterojunctions on the fundamental design parameters of nanophotonic devices," *Journal of Nanophotonics*, vol. 8, no. 1, pp. 083 096–083 096, 2014.

[10] P. Harrison and A. Valavanis, *Quantum Wells, Wires and Dots*. John Wiley & Sons, Inc., 2016.

[11] I. Tan, G. L. Snider, L. D. Chang, and E. L. Hu, "A self-consistent solution of Schrödinger–Poisson equations using a nonuniform mesh," *Journal of Applied Physics*, vol. 68, no. 8, pp. 4071–4076, 1990.

[12] M. Claus, S. Mothes, S. Blawid, and M. Schröter, "COOS: a wave-function based Schrödinger–Poisson solver for ballistic nanotube transistors," *Journal of Computational Electronics*, vol. 13, no. 3, pp. 689 – 700, 06 2014.

[13] G. Hałdaś, "Implementation of non-uniform mesh in non-equilibrium Green's function simulations of quantum cascade lasers," *Journal of Computational Electronics*, vol. 18, no. 4, pp. 1400–1406, 2019.

[14] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge University Press, 2005.

[15] S. Blawid, M. Claus, and M. Schroter, "Phenomenological Modeling of Charge Injection - Beyond the Schottky Barrier Paradigm," *ECS Transactions*, vol. 49, no. 1, pp. 85 – 92, 10 2012.