Atomistic quantum transport approach to time-resolved device simulations

Bozidar Novakovic* and Gerhard Klimeck†
School of Electrical and Computer Engineering and Network for Computational Nanotechnology
Purdue University, West Lafayette, Indiana 47907
Email: *novakovic@purdue.edu, †gekco@purdue.edu

***Abstract*—Having access to time-resolved quantum transport data is beneficial for more accurate calculation of energy/delay device characteristics during turn on, for studying novel effects based on the wave function phase manipulation, and as an alternative research path to simulating dissipation and nonlocal scattering in real time. We present a time-resolved version of the quantum transmitting boundary method that relies on the efficient algorithms developed previously for the steady state version. Our method in principle can handle arbitrary time-dependent bias at gate and current-carrying lead terminals, where leads are limited to rigid spatial potential, and arbitrary atomistic geometries in the semi-empirical tight-binding basis. Using our method in the wide-band approximation, therefore relaxing the numerical complexities of energy scattering, we present the time-resolved results for important device quantities and discuss the limitations of the wide band approximation. We also discuss the potential of this method for parallelization by showing the computation time versus number of processes scaling results for multiple levels of parallelization.**

I. Introduction

When predicting the performance of future nanoelectronic devices, one of the most important parameters is the energy-delay characteristic [1]. It shows how much energy is necessary to turn the device on and how long does it take to do it. In the absence of time-resolved data one has to resort to steady state IV characteristics and capacitances derived from them in order to calculate the turn on energy and delay. With our atomistic time-resolved quantum transport approach we attempt to introduce methodology that will improve the accuracy of energy-delay predictions by having access to time-resolved current, density, and other electronic quantities. Some other interesting application areas for this time-resolved quantum transport method are studying the effects of the wave function phase manipulation [2] and exploitation of the spatio-temporal features of the self-consistent device potential [3]. Furthermore, our method can be viewed as the first step in creating methods for simulating device dissipation due to phonon scattering using real time approach that may favourably compare to traditional approaches in certain application areas.

There are multiple advantages of the time resolved quantum transport method presented in this paper. First, as part of our comprehensive simulation tool [4] it can handle realistic device geometries and materials in atomistic semi-empirical tight-binding basis. Another related benefit is the utilization of the existing code infrastructure for the time-resolved self-consistent calculations via the nonlinear Poisson equation at frequencies where the quasi-electrostatic approximation holds. Furthermore, for not too fast excitations our method directly relies on the efficient transfer matrix method (TMM) derived for steady state simulations to calculate the lead modes and self-energies $\Sigma(\mathrm{E})$ [5]. Although the TMM comes with a limitation at the same time, namely the leads must be periodic, this still enables using our method for the majority of device applications commonly encountered, except the devices with disordered leads and driving frequencies in the vicinity of the plasma frequencies. Specifically, we assume that the leads have a time-dependent spatially rigid potential, which is a common approximation in device simulations due to high plasma frequencies of doped semiconductors (e.g. sec. 13.2 in ref. [6]).

II. Method

Our method is based on the mixed energy-time space approach with the scattering matrix, where the wave function is partitioned as a sum of injected and scattered terms. The injection basis consists of single energy Bloch waves, while the scattered waves in general can contain multiple energies due to the time-dependent device potential. Being the time-resolved version of the quantum transmitting boundary method (QTBM) [7], [8], [5], this method has clear connection to Greens functions approach [9]. Since it seems that QTBM may not be a universally accepted term for the kind of method described in refs. [7], [8], [5] and in this work, we explicitly define QTBM to be a method that yields a boundary value problem through explicit incorporation of the open boundary conditions. This differs from the initial value problem obtained by using the transfer matrix method, where the solution is assumed (up to a constant) on one side of the device and then integrated through to the other side.

We start from the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\psi\left(t\right)}{\partial t}=H\left(t\right)\psi\left(t\right)\ , \tag{1}$$

and the periodic lead assumption, where the lead wave functions are expressed in terms of forward and backward travelling Bloch waves

$$\begin{aligned}\Psi_{0}\left(t\right)&=\Phi^{+}\mathrm{A}\left(t\right)+\Phi^{-}\mathrm{B}\left(t\right)\ ,\\ \Psi_{-1}\left(t\right)&=\Phi^{+}\mathrm{A}\left(t\right)\left[\Lambda^{+}\right]^{-1}+\Phi^{-}\mathrm{B}\left(t\right)\left[\Lambda^{-}\right]^{-1}\ ,\end{aligned} \tag{2}$$

where 0 and -1 denote two adjacent slabs in the lead, $\Phi^{\pm}$ are the forward and backward propagating lead wave functions in slab 0, and $\Lambda^{\pm}$ are the phase factors between the two slabs. Slab is a region in the lead with a single unit cell thickness in the transport direction, with the other two dimensions being determined by the device transversal boundary conditions. B is

978-1-4673-7860-4/15/$31.00 ©2015 IEEE

the scattering coefficient to be calculated, and A is the injection condition. The injection condition for the time-dependent rigid lead is given by:

$$\mathrm{A}\left(t\right)=\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t}\left(E+V(t)\right)dt\right]\mathbf{1}\ , \tag{3}$$

where E is the total energy of the injected lead state and V(t) is the change in the lead potential energy due to external time-dependent bias. After some algebra, the main equation for the time-resolved wave function with open boundaries is:

$$\mathrm{i}\hbar\partial\Psi_{\mathrm{E}}\partial\mathrm{t}=[\mathrm{H}_{\mathrm{E}}(\mathrm{t})-\mathrm{E}(\mathrm{t})]\Psi_{\mathrm{E}}(\mathrm{t})+\mathrm{I}_{\mathrm{E}}(\mathrm{t})\ , \tag{4}$$

where Ψ_{E}, h_{E} , and I_{E} are the wave function, Hamiltonian, and lead injection terms in the mixed scattered mode/orbital space. An equivalent equation in just orbital space can be obtained by explicitly calculating $\Sigma(\mathrm{E})$ and summing it with the orbital Hamiltonian.

When the above equation is used directly with the lead TMM, the phase coherence between the device and leads may be lost due to the arbitrary phase of the TMM eigensolver solution. This can be resolved by consistent phase normalization in the leads. Another important computational aspect is related to the scattered-mode degrees of freedom in h_{E} that belong to multiple closely spaced energies due to the time-dependent nature of the excitation. Naive approach may lead to a large matrix condition number and numerical instabilities when extracting scattering probabilities to different energies. However, for slow enough excitations one can use the wide-band approximation (WBA), which amounts to keeping only the injected energy in the scattered-mode space, or equivalently replacing the convolution $\int \mathrm{du}\Sigma(\mathrm{t}-\mathrm{u})\Psi_{\mathrm{E}}(\mathrm{u})$ with $\Sigma(\mathrm{E})\Psi_{\mathrm{E}}$, which significantly reduces the computational burden.

III. Results

Here, we show the time-resolved data using the WBA, as well as the domain of validity of the WBA in terms of the rate of excitation. The simulated nanostructure is a 3x3 unit cell Si nanowire $20\mathrm{nm}$ long in $\mathrm{sp3d5s*}$ atomistic tight binding basis, Fig. 1. The time dependent potential in the middle of the nanowire is a $4\mathrm{nm}$ wide potential barrier with $10\mathrm{mV}$ amplitude and variable rise times. Figs. 2 and 3 show the time-resolved wave function magnitude, transmission T, and reflection R probabilities for energy with one propagating mode. The wave function magnitude is plotted along a chain of atoms with cross-sectional coordinates $1.4\mathrm{x}0.3\mathrm{nm}$ and shows artifact free lead/device interface, i.e. without reflections usually associated to non-ideal open boundary conditions. Fig. 4 shows the domain of validity of the WBA based on the probability current conservation. For all three excitations rates the final potential barrier amplitude is $10\mathrm{mV}$. The faster the excitation the less conserved the probability current is in the WBA due to the fact that faster excitations correspond to larger energy changes of the initial injected state. The final wave function magnitudes match quite well for the three excitation rates. However, from the probability conservation we conclude that the WBA is valid up to around $1\mathrm{mV/ps}$, which corresponds to roughly a few GHz in realistic devices.

Nanowire calculations mentioned in the previos parahraph have the potential to be parallelized in order to be able to speed

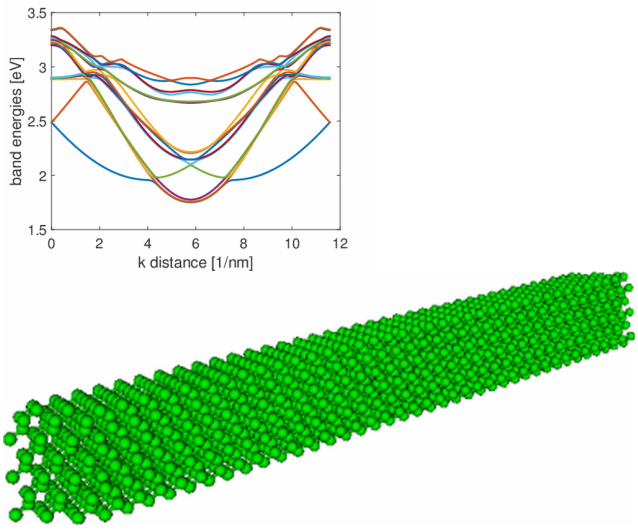

Fig. 1. Simulation structure. Bottom: Si nanowire 3x3 unit cells cross section, $20\mathrm{nm}$ length. Top: $\mathrm{sp3d5s*}$ band diagram where horizontal axis is the longitudinal wave vector in terms of its distance from the far left hand side of the Brillouin zone.

up realistic device calculations with large number of relevant energies and bigger nanowire cross sections. We employ two levels of parallelization: the first level is the parallelization of the energy space and the second level is the parallelization of the linear system itself (eq. 4) at each energy. Fig. 5 shows the surface plot of transmission as a function of time and energy with 64 total energies and 20000 time steps, where the energies are parallelized across 64 processes and the linear system across 4 processes, utilizing in total 256 computing cores. Fig. 5 uses the same device and parameters as fig. 3. It shows abrupt transmission changes along the vertical axis at $\mathrm{t}=0\mathrm{ps}$, corresponding to the onset of different bands in the lead, and then gradual decrease of transmission up to around $\mathrm{t}=10\mathrm{ps}$ corresponding to the increase in the tunneling barrier amplitude. The energy windows where the transmission drops roughly correspond to the steady barrier amplitude.

Finally, figs. 6 and 7 show the scaling performance for both levels of parallelization for a Si nanowire with $4\mathrm{x}4$ unit cells $(2.15\mathrm{x}2.15\mathrm{nm})$ and 10 time steps. In fig. 6 there are 64 energy levels. The linear system (time stepper) parallelization is fixed to 4 ranks, and the total number of ranks is varied, thus showing the effect of energy parallelization. The scaling is almost ideal, which is not surprising for an optimized code due to different energies being independent. In fig. 7 there is only one energy, while the time stepper parallelization is varied, thus showing the performance of the linear system solution in parallel. The linear system solver is direct, based on the LU decomposition (MUMPS LU). The scaling is good, albeit not ideal. The improvements may be possible by further optimizing the direct solver and also by investigating the possibility of using iterative linear solvers.

IV. Summary and Conclusion

We present a time-resolved quantum transmitting boundary method suitable for atomistic device simulations in the

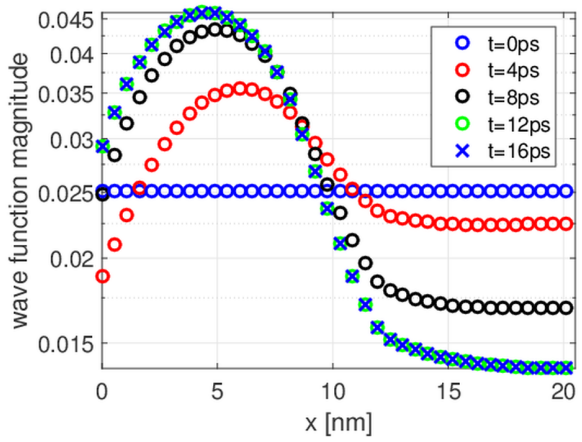

Fig. 2. Wave function magnitude along a chain of atoms at 1.4x0.3nm for left injection at energy 1.755eV with one propagating mode. The rising edge of the potential barrier in the middle is 1mV/ps and the total rise time is 10ps. The time step is 1fs. There are no visible artificial reflections at device/lead interfaces.

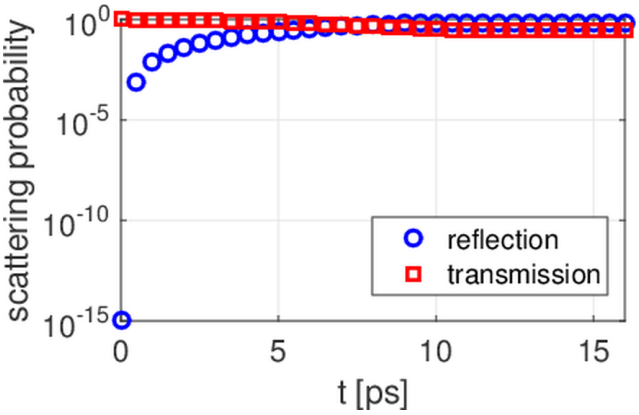

Fig. 3. T and R corresponding to data in Fig. 2.

quantum transport framework and semi-empirical tight-binding basis. We show some basic time-resolved device quantities that can be obtained with this method and from which observable quantities, like density and current, can be straightforwardly calculated. We employ the wide band approximation and discuss its physical limitations using a Si nanowire model device. Finally, we present the parallelization performance of the algorithm and its implementation in terms of the computational time scaling with the number of computational cores. Overall, our method shows a good potential to simulate realistic devices in the quantum transport regime and to successfully expand the current steady state quantum transport simulation methodologies to time-resolved domain.

ACKNOWLEDGMENT

The authors would like to acknowledge the funding support from the U.S. Army Research Office contract No. W911NF-12-1-0607 and W911NF-08-1-0527, and the usage of nanoHUB.org computational resources funded by the NSF grant No. EEC-1227110. B.N. would like to thank Harshad

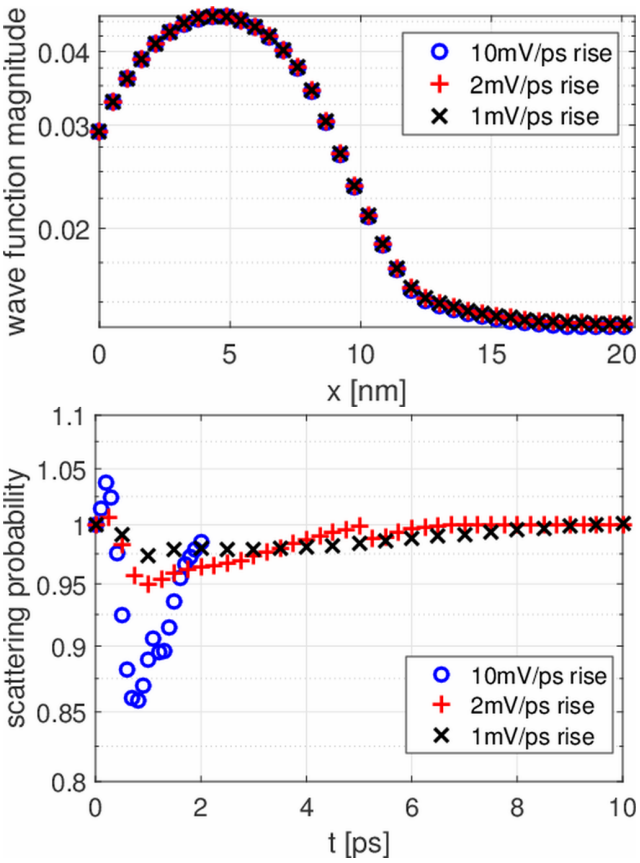

Fig. 4. Long time wave function magnitude and probability current conservation (T + R) for 1, 2, and 10mV/ps rising edges. Due to non-ideal probability current conservation the wide-band approximation is valid up to around 1mV/ps, corresponding to a few GHz in realistic transistors.

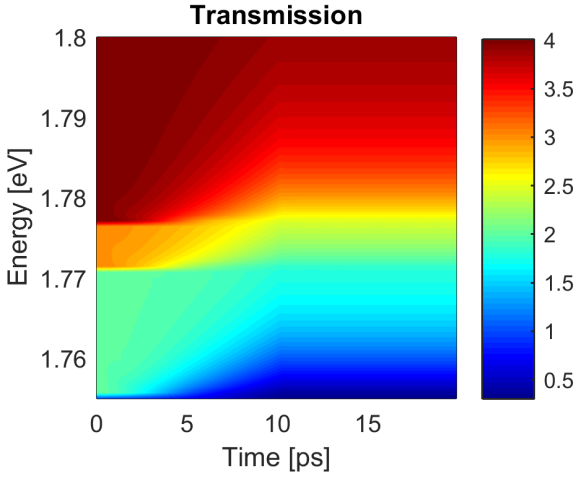

Fig. 5. Transmission as a function of time and energy with 64 total energies and 20000 time steps. Same parameters as in fig. 3. Abrupt transmission changes along the vertical axis at t = 0ps correspond to the onsets of different bands in the lead, while the gradual decrease of transmission up to around t = 10ps corresponds to the increase in the tunneling barrier amplitude. The energy windows where the transmission drops roughly correspond to the steady state barrier amplitude.

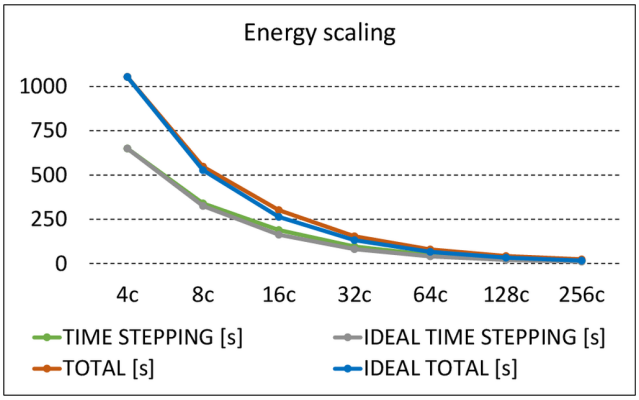

Fig. 6. Strong scaling study for a Si nanowire with 4x4 unit cell cross section and 10 time steps. The time stepper (linear system) parallelization is fixed to 4 ranks. The total number of ranks is varied, thus varying the energy parallelization. The scaling is almost ideal, which is expected for an optimized code due to different energies being independent.

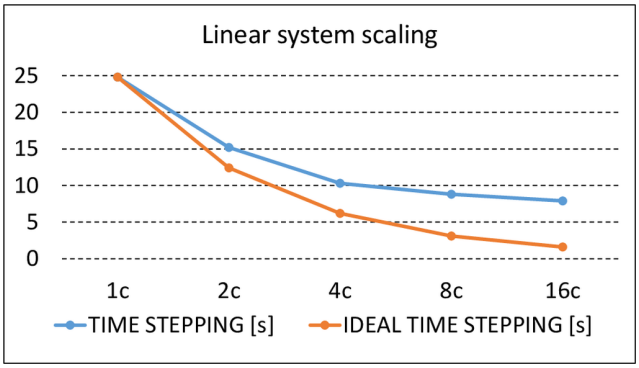

Fig. 7. Same as fig. 6, except that there is only 1 energy, while the time stepper (linear system) parallelization is varied. The linear solver is MUMPS direct solver based on the LU decomposition. The scaling is good, even if not ideal. Further improvements may be possible by optimizing the direct linear solver and by investigating the possibility of using iterative linear solvers.

Sahasrabudhe for useful discussions and help with some parts of the code.

REFERENCES

[1] D. E. Nikonov and I. A. Young, *IEEE Journal on Exploratory Sold-State Computational Devices and Circuits*, pp. 1–8, 2015, doi: 10.1109/JXCDC.2015.2418033.

[2] B. Gaury and X. Waintal, *Nat. Commun.*, vol. 5, p. 3844, 2014.

[3] G. Stefanucci, S. Kurth, A. Rubio, and E. K. U. Gross, *Phys. Rev. B*, vol. 77, p. 075339, 2008.

[4] J. E. Fonseca, T. Kubis, M. Povolotskyi, B. Novakovic, A. Ajoy, G. Hegde, H. Ilatikhameneh, Z. Jiang, P. Sengupta, Y. Tan, and G. Klimeck, *J. Comput. Electron.*, vol. 12, pp. 592–600, 2013.

[5] M. Luisier, A. Schenk, and W. Fichtner, *Phys. Rev. B*, vol. 74, p. 205323, 2006.

[6] H. Haug and A.-P. Jauho, *Quantum Kinetics in transport and optics of semiconductors*. Springer, Berlin, 1998.

[7] C. S. Lent and D. J. Kirkner, *J. Appl. Phys.*, vol. 67, p. 6353, 1990.

[8] S. E. Laux, A. Kumar, and M. V. Fischetti, "Analyisis of quantum ballistic electron transport in ultrasmall silicon devices including space-charge and geometric effects," *J. Appl. Phys.*, vol. 95, pp. 5545–5582, 2004.

[9] B. Gaury, J. Weston, M. Santin, M. Houzet, C. Groth, and X. Waintal, *Phys. Rep.*, vol. 534, pp. 1–37, 2014.