Linear scaling techniques for first-principle calculations of large nanowire devices

D. Zhang and E. Polizzi

Department of Electrical and Computer Engineering
University of Massachusetts, 01003 Amherst, USA
dzhang@ecs.umass.edu, polizzi@ecs.umass.edu

ABSTRACT

The purpose of this paper is to improve performances for the calculation of the electron density at a given step of the self-consistent first-principle procedure (for one given atomistic potential configuration). We propose to use a combination of numerical techniques and demonstrate their robustness and scalability for simulating a 3D Carbon nanotube (CNT) device.

Keywords: First-principle calculations, DFT/Kohn-Sham, Carbon nanotube, FEM, Electronic structure, Mode approach, Contour integration

1 Introduction

In nanoelectronics, large scale 'ab-initio' simulations could significantly enhance our understanding of nanoscale physics and engineering related issues of materials and transistor devices. At large scale (scale of a transistor), first-principle atomistic simulations of devices are in need of appropriate efficient modeling strategies and innovative numerical algorithms.

In this study, we propose to make use of a full 3D atomistic real-space mesh technique to discretize the DFT/Kohn-Sham equations for an entire CNT device. Real-space mesh techniques for first-principle electronic and transport calculations (finite element method- FEM- in our case) can exhibit significant numerical advantages as compared to other traditional discretization techniques such as plane waves expansion schemes or linear combination of atomic orbitals [1]. However, they have been so far limited to the simulation of small devices containing only few number of atoms. One limiting factor is concerned with the high numerical costs that are required for calculating the electron density at each step of the DFT/Kohn-Sham self-consistent iterations. In [2], we have presented an effective combination of the following three $O(n)$ numerical techniques for addressing these problems for nanowire-type devices: (i) the electron density is calculated by performing a contour integration of the Green's function along the complex energy plane [3], [4], (ii) a mode approach is used to reduce the size of the discretized problem (typically from $O(10^8)$ to $O(10^5)$) while producing very narrow banded matrices [5], (iii) the diagonal elements of the Green's function are obtained using our in-house $O(n)$ banded system solver. The summary chart of our numerical approach is showed in Figure 1.

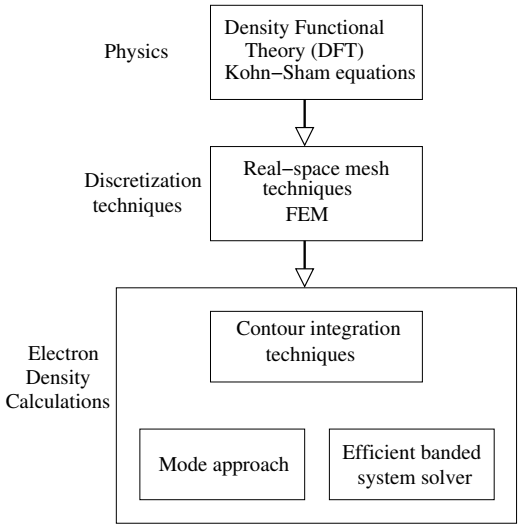

Figure 1: Summary chart of our combined modeling techniques for computing the electron density using first principle calculations.

Our results have shown that the time needed to compute the electron density in electronic structure calculations using our contour integration technique, is now order of magnitude faster than the one obtained using other traditional approaches such as solving eigenpairs. Here, we propose to discuss how we can improve the robustness and scalability of those three techniques and their combination while performing numerical simulations on 3D Carbon nanotube (CNT).

2 Numerical scaling techniques for electronic density calculations

2.1 The contour integration technique

For isolated systems, the contour integration technique allows solving a large number of eigenpairs of the Kohn-Sham equation to be replaced by computing the diagonal elements of few Green's function matrices ($O(10)$) in the complex plane[3]. The expression of the electron density is then given by:

NSTI-Nanotech 2008, www.nsti.org, ISBN 978-1-4200-8503-7 Vol. 1

$$n(\mathbf{r}) = -\frac{1}{\pi}\left(\int_C dz \mathrm{Im}(G(\mathbf{r},\mathbf{r},Z_E))f_{FD}(Z_E - E_F)\right) - \frac{1}{\pi}\sum_{poles,n} \mathrm{Im}(-2i\pi k_B T G(\mathbf{r},\mathbf{r},Z_{E_n})) \quad (1)$$

where $G(\mathbf{r},\mathbf{r},Z_E)$ represents the diagonal elements of the Green's function for the complex energy Z_E, the complex contour C contains all the resonances on the real axis (eigenvalue solutions of the Kohn-Sham equation), and the second term of the right hand side corresponds to the contribution of the residues of all poles on the imaginary axis that are introduced by the Fermi-Dirac distribution function f_{FD}. One should note that $Z_{E_n} = E_F + i(2n+1)\pi k_B T$. For open systems (transport problem), this approach can be used as well to perform the integration on the equilibrium part of the electron density while avoiding solving a prohibitive number of linear systems[4].

Since, in practice, only the diagonal elements of few Green's functions need to be calculated along the complex plane (this number is independent of the size of the system), the contour integration technique represents an attractive alternative approach to the traditional eigenvalue problem. In order to take fully advantage of the contour integration technique, however, we need to overcome the major numerical difficulties in inverting directly the Hamiltonian matrix (to obtain the Green's function) that can reach size of millions (for large atomistic systems). For matrices that are banded, which are naturally obtained after discretization with nanotubes or any type of nanowire devices, it is indeed possible to make use of an efficient algorithm to compute only the diagonal elements of the inverse of a banded matrix. This algorithm, which will be presented in section 2.3, exhibits a linear scaling performance $O(b^2 n)$ with the size of the system n and a quadratic scaling with the bandwidth b. Therefore this scheme, and then the contour integration technique, is particularly well adapted for narrow banded systems. In order to reduce the bandwidth of the Hamiltonian matrix obtained by the finite element method (FEM), we make use in the following of a subband decomposition technique (mode approach) [5].

2.2 The mode approach

One of the underlying advantage of using an atomistic real-space mesh discretization, is to be able to rigorously perform a mode approach where the quantum confinement and transport problems are treated separately. The $3D$ wave-function are expanded by

$$\Psi_{3D}(x,y,z) = \sum_{n=1}^{\infty} \psi_n(x)\chi_n(y,z), \quad (2)$$

where y, z are the coordinates in the cross section of the wire. and the calculation procedure is as follows:
(a) Solving only one $2D$ Schrödinger equation with closed (Dirichlet) boundary condition to obtain the eigenvalues E_n and corresponding eigenfunctions $\chi_n(y,z)$ in the cross section:

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\chi_n(y,z) + U_{2D}\chi_n(y,z) = E_n\chi_n(y,z). \quad (3)$$

The same mathematical complete basis set $\{\chi_n(y,z)\}$ is then used for all cross sections along the nanotube ($\forall x$). Therefore, a suitable choice for the 2D potential U_{2D} can be defined by [5], [6]

$$U_{2D}(y,z) = \frac{\int_x U(x,y,z)n(x,y,z)}{\int_x n(x,y,z)} \quad (4)$$

where the regions "seen by the electrons" are captured in average along the longitudinal x direction.
(b) Solving the following 1D coupled Schrödinger equation (with appropriate boundary conditions):

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi_n(x) + \sum_{m=1}^{\infty} U_{mn}\psi_n(x) = (E - E_n)\psi_n(x) \quad (5)$$

$$U_{mn}(x) = \int_{y,z} (U - U_{2D})\chi_n\chi_m dydz \quad (6)$$

When this decomposition is accounting for all coupling between modes and with a sufficient number of modes, the mode approach is equivalent to obtaining the full 3D real-space solutions. If only M modes are taken into account in the calculations, the size of this system matrix and bandwidth can be both reduced by a factor $N_{Y,Z}/M$ (where $N_{Y,Z}$ is the number of nodes in the cross section and $M << N_{Y,Z}$ in practice). Using an appropriate reordering, the obtained matrix is block tridiagonal, each $M \times M$ blocks being completely dense. Using this approach, however, all the diagonal blocks of size $M \times M$ of the obtained mode approach Green's function matrix g are needed in the calculation of the density since:

$$G(\mathbf{r},\mathbf{r},Z_E) = \sum_{m,n} \chi_m(y,z)\chi_n(y,z)g_{m,n}(x,x,Z_E).$$

where $g_{m,n}(x,x',Z_E)$ denotes the matrix elements of g.

2.3 Banded system solver

Since the obtained mode approach matrices are narrow banded and block tridiagonal (as described above), in Figure 2 we show that it is possible to make use of a LU factorization and modified forward and backward sweeps which focus only in computing the diagonal blocks of the Green's function. This banded system solver for diagonal elements is the numerical algorithm underneath the recursive Green's function technique[7]. The computational cost of this technique is

$O(b^2N)$ which is linear in N , and takes advantage of the small bandwidth b obtained using the mode approach. In addition, our diagonal system solver has been implemented using LAPACK and BLAS level 3 routines to minimize memory references and increase performances.

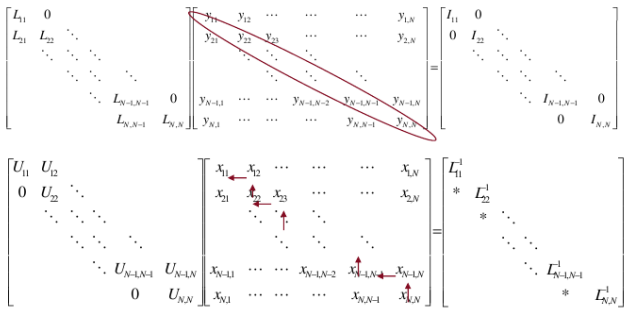

Figure 2: Overview of the different solve steps for obtaining the diagonal blocks of the inverse of a symmetric and non singular banded matrix after a LU factorization without pivoting (i.e. obtaining the diagonal blocks of X where AX=I and A=LU). For the forward sweep $LY = I$, we compute only the diagonal blocks of Y which corresponds to the inverse of each local diagonal blocks of L. Then for the backward sweep we can proceed as follows: (i) solve $U_{n,n}X_{n,n} = L^{-1}_{n,n}$ to obtain $X_{n,n}$ (ii) solve $U_{n-1,n-1}X_{n-1,n} = -U_{n-1,n}X_{n,n}$ to obtain $X_{n-1,n}$, (iii) since the matrix is symmetric it comes $X_{n,n-1} = X_{n-1,n}$, (iv) Solve $U_{n-1,n-1}X_{n-1,n-1} = L^{-1}_{n-1,n-1} - U_{n-1,n}X_{n,n-1}$ to obtain $X_{n-1,n-1}$, (v) repeat recursively from step (ii) with $n <= n-1$ until $n = 1$.

3 Results

In order to illustrate the efficiency of the above techniques, one needs to consider two parameters: M, the number of modes that will be considered, and N_E, the number of points of discretization for the complex energy contour. For a given atomistic potential configuration (the Hartree and exchange-correlation potentials are given), we propose to compute the electron density of an isolated (13,0) CNT with arbitrary length by solving the Schrödinger-type equation with different values for M and N_E.

3.1 Validity of the contour integration technique

In table 1, we propose first to fix the number of modes M and compare the results obtained on the electron density with the contour integration technique and different N_E, with the ones obtained with a traditional eigenvalue solver (used as reference). The results indicate that using an uniform repartition of the energy points along the contour, the poles Z_{E_n}, which appear in equation (1) at room temperature, may not be captured appropriately. This gives rise to inconsistency in the error while varying the number of energy points N_E. In order to obtain an identical accuracy at both zero and room temperatures, we propose in Figure 3 to make use of a non-uniform repartition of energy points allowing a much larger concentration of contour points around $Re(Z_{E_n})$.

N_E	1-unit ($T = 0K$)	1-unit, uniform ($T = 300K$)	1-unit, non-uniform ($T = 300K$)
40	0.85%	1.3%	0.50%
60	0.34%	7.1%	0.34%
80	0.19%	0.6%	0.22%
100	0.14%	1.5%	0.22%
200	0.12%	0.3%	0.18%

Table 1: Relative residual (error) $||n_e - n_c||_\infty / ||n_e||_\infty$ for different calculations of the electron density in function of N_E by two techniques: eigenvalue problem (n_e) and contour integration (n_c), for 1-unit CNT with 50 modes included. A non-uniform repartition of energy points becomes necessary at room temperature to capture the abrupt variation of the Green's function caused by the poles Z_{E_n}.

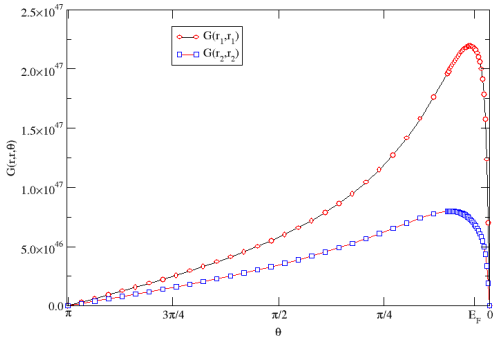

Figure 3: Variation of the Green's function at a given position in space along the complex contour of integration going from $\theta = 0$ to $\theta = \pi$ (half-circle), and using a non-uniform repartition of the energy points. For the purpose of illustrating a simple case, we select two particular real space points $\mathbf{r_1}$ and $\mathbf{r_2}$ inside the CNT, one at the position of a center of an atom and the other one between two atoms. The contour integration of these two curves give the electron density at position $\mathbf{r_1}$ and $\mathbf{r_2}$. Using a non-uniform mesh (here 60 points), one can see that the abrupt variation of the Green's function is well captured, and in table 1 we obtain a 0.34% error accuracy as compared to solving the eigenvalue problem.

3.2 Validity of the mode approach

In order to illustrate the robustness of the mode approach, we propose to fix the number of energy points along the contour integration to $N_E = 100$ and vary the

NSTI-Nanotech 2008, www.nsti.org, ISBN 978-1-4200-8503-7 Vol. 1

number of modes from $M = 1$ to $M = 100$. In Figure 4, the case $M = 100$ is also used as the reference solution to calculate the error, since for this large number of modes and coupling between modes the mode approach and the full 3D solutions can be considered equivalent. The curves show that the error decays exponentially as we increase the number of modes and for $M = 50$, for example, the error is already below 2%.

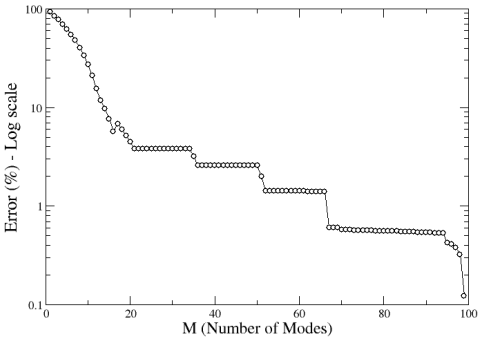

Figure 4: Error on the electron density obtained using different number of modes M and compared to a reference solution where $M = 100$.

3.2.1 Scalability of the proposed techniques

Table 2 summarizes the time obtained by the combination of our proposed technique for computing the electron density of CNT with different lengths. One can

#unit cells	1-unit	3-unit	6-unit	24-unit	48-unit
$L(nm)$	0.57	1.42	2.70	10.37	20.59
$\#atoms$	52	156	312	1248	2498
$T_{eigen}(s)$	283	5806	N/A	N/A	N/A
$T_{contour}(s)$ $N_E = 100$	33	86	165	629	1245
$T_{contour}(s)$ $N_E = 60$	20	52	100	381	754

Table 2: Times obtained using our contour integration technique ($N_E = 60$ and 100, $M = 50$) and solving the eigenvalue problem for obtaining the electron density of a (13,0) CNT up to 48-unit cell (2498 atoms). The calculations are performed on only one core of a Clovertown $2.66Ghz$ and $16Gb$. The eigenvalue problem is solved using the direct banded eigenvalue solver in LAPACK and exhibits memory limitations beyond 3-unit cells.

see that the time for the contour integration technique scales linearly with the length of the device, and it is directly proportional to the number of energy points in the contour N_E. If we compare these results with those obtained with a more traditional eigenvalue solver technique (which takes also advantage of the mode approach), one observe a $\sim \times 100$ speed up improvement for 3-unit cells.

4 Discussions

The proposed numerical scheme is applicable to any nanowire type device independently of their size (as long as the cross section of the wire is much smaller than its length). While increasing the length of the device, we have demonstrated in Table 2 that the computational cost of the our numerical scheme is linear. The efficiency of this technique will eventually exhibit limitations if the cross section of the nanowire does not contain less than $\sim$ 100 atoms (in order to minimize the number of modes that are needed). It would be possible, however, to address these limitations by coupling the proposed approach with an outer layer domain decomposition technique.

It is also possible to perform transport calculations within the framework of the proposed scheme. To account for the flow of electrons, small dense blocks (self-energy matrices) need to be added at the edges of the the obtained block tridiagonal Hamiltonian matrix with the mode approach. Since all the blocks are already treated as dense blocks, it is possible to solve the resulting linear systems (resulting from the NEGF transport equation [8]) with at least the same efficiency than for the isolated systems (the efficiency will be in practice much higher since the diagonal elements are not needed here and the complexity of the forward and backward sweeps will be reduced).

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CCF 0635196 and EEC 0725613.

REFERENCES

[1] T. Beck, Rev. of Modern Phys., 72, 1041, (2000).
[2] D. Zhang, E. Polizzi, Journal of Comput. Electronics, to appear(2008).
[3] R. Zeller, J. Deutz, and P. H. Dederichs, Solid State Commun. 44, 993-997 (1982).
[4] M. Brandbyge, J.-L. Mozos, P. Ordejn, J. Taylor, and K. Stokbro, Phys. Rev. B 65, 165401 (2002).
[5] E. Polizzi, N. Ben Abdallah, J. of Comput. Phys., 202, 150-180 (2005).
[6] E. Polizzi, N. Ben Abdallah Phys. Rev. B 66, 245301 (2002).
[7] F. Sols, M. Macucci, U. Ravaioli, J. Appl. Phys. 66 (8), 3892 (1989).
[8] E. Polizzi, A. Sameh, H. Sun, 2004 NSTI Nanotechnology Conference and Trade Show. Technical Proceedings, Vol. 2, Chapter 8, pp403-406 (2004)