

10-20-2008

# Contact Block Reduction Method for Ballistic Quantum Transport with Semi-empirical $sp^3d^5s^*$ Tight Binding band models

Hoon Ryu

*Purdue University - Main Campus*

Gerhard Klimeck

*Purdue University - Main Campus, gekco@purdue.edu*

Follow this and additional works at: <http://docs.lib.purdue.edu/nanodocs>

---

Ryu, Hoon and Klimeck, Gerhard, "Contact Block Reduction Method for Ballistic Quantum Transport with Semi-empirical  $sp^3d^5s^*$  Tight Binding band models" (2008). *Other Nanotechnology Publications*. Paper 147.  
<http://docs.lib.purdue.edu/nanodocs/147>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# Contact Block Reduction Method for Ballistic Quantum Transport with Semi-empirical $sp^3d^5s^*$ Tight Binding Band Models.

Hoon Ryu<sup>1\*</sup> and Gerhard Klimeck<sup>1</sup>

<sup>1</sup>Network for Computational Nanotechnology, Birck Nanotechnology Center, Purdue University, West Lafayette, IN 47906, USA.

\*Email: ryu2@purdue.edu

## Abstract

The utility of the *Contact Block Reduction* method (CBR) to find the retarded Green's function for ballistic quantum devices with semi-empirical tight binding band (TB) models is discussed. This work shows that the original method needs several modifications to be used with TB models. In the common case where two contacts are used for transport in quantum wires, our approach computes the transmission coefficients with much less computing load than the state-of-the-art Recursive Green's Function (RGF) algorithm.

## 1. Motivation

CBR method, first suggested by Mamaluy *et al.*,<sup>1</sup> has received attention in recent years due to its ability to compute the retarded Green's function for open systems with low computing intensity. Its successful application, however, was shown only with the effective mass band (EM) model in cubic-grid bases.<sup>2,3</sup>

The EM model works well near the conduction band minima for large devices. The quantized states, however, are not accurate if devices are on the nm-scale.<sup>4</sup> For the correct modeling of nano-scale device behaviors, one therefore should use more sophisticated band models with an atomistic basis representation, which reproduce experimentally verified band structure of semiconductor crystals.<sup>5</sup> This, however, requires much larger computing expense.

The CBR method coupled with the most sophisticated band model, therefore, may provide an excellent utility since both quantum and atomistic effects can be properly with reasonable compute requirements. Throughout this work, we use the semi-empirical  $sp^3d^5s^*$  TB band model with a set of zincblende-local (ZB) orbital bases, which has shown its accuracy in estimating band structures of various nano-scale devices and semiconductor crystals.<sup>5,7</sup>

## 2. Methodology

A real device is coupled with contacts to allow carrier in-and-out flow, forming an open system described with a non-Hermitian system matrix in the non-equilibrium Green's function (NEGF) formalism,<sup>8</sup> with which we can model any quantum devices and estimate their I-V and

charge profiles by computing transmission functions and local density of states. The evaluation of these quantities requires the retarded Green's function  $G^R(E)$  for an open system as a function of energy  $E$  as defined in (1), where  $H^o$  is the Hamiltonian of the closed device and  $\Sigma$  is the complex self-energy term which expresses the coupling between contacts and the device.

$$G^R(E) = [(E + i\eta)I - H^o - \Sigma(E)]^{-1}, \quad \eta \rightarrow 0^+ \quad (1)$$

For 3D open systems, a huge computation is required to find the partial inverse of the system matrix in (1). RGF significantly reduces this numerical load,<sup>9,10</sup> however, it is still expensive and more efficient alternatives need to be considered,<sup>11</sup> one of which is CBR since it reduces the matrix inversions to find  $G^R$ .

### 2.1 CBR with Tight Binding Parameters

The first step in CBR is to divide the device space into the boundary region  $c$  that couples with the contacts, and region  $d$  for the rest. Since the self-energy  $\Sigma$  is non-zero only in the region  $c$ ,  $H^o$  and  $\Sigma$  can be decomposed as (2), where the subscripts  $\{c, d\}$  denote corresponding regions. Then  $G^R$  can be computed with the Dyson equation in (3), where we conditioned  $\Sigma^x$  and  $G^x$  with a Hermitian matrix  $X$  to minimize matrix inversions as shown in (4).

$$H^o = \begin{bmatrix} H_c^o & H_{cd}^o \\ H_{dc}^o & H_d^o \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_c & 0_{cd} \\ 0_{dc} & 0_d \end{bmatrix} \quad (2)$$

$$A_c^{-1} = (I_c - G_c^x \Sigma_c^x)^{-1}, \quad (3)$$

$$G^R(E) = (I - \Sigma^x G^x)^{-1} G^x$$

$$= \begin{bmatrix} A_c^{-1} & 0_{cd} \\ -G_{dc}^x \Sigma_c^x A_c^{-1} & I_d \end{bmatrix} \begin{bmatrix} G_c^x & G_{cd}^x \\ G_{dc}^x & G_d^x \end{bmatrix}$$

$$X = \begin{bmatrix} x_c & 0_{cd} \\ 0_{dc} & 0_d \end{bmatrix}, \quad \Sigma^x = \Sigma - X, \quad (4)$$

$$G^x = [EI - (H^o + X)]^{-1}$$

$$= \begin{bmatrix} G_c^x & G_{cd}^x \\ G_{dc}^x & G_d^x \end{bmatrix} = \sum_{\alpha} \frac{|\Psi_{\alpha}\rangle \langle \Psi_{\alpha}|}{E - \varepsilon_{\alpha} + i\eta} \quad (\eta \rightarrow 0^+),$$

where  $\varepsilon_{\alpha}, \Psi_{\alpha}$  are the eigenvalues/vectors of  $(H^o + X)$

Here, matrix inversions may not be a problem unless devices have very large boundary region, since one only needs to invert a boundary block  $A_c$ . Therefore, the major task becomes to solve eigenstates for a Hermitian matrix ( $H^o+X$ ). The true speed-up, however, is never available without a significant reduction in the number of used eigestates, which can be achieved via a smart choice of  $X$  to make  $\Sigma^x$  negligible in the energy range of interest.<sup>2</sup>

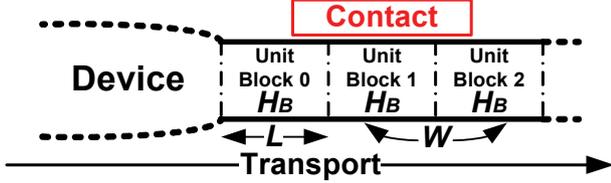


Figure 1. Schematic picture of the semi-infinite contact.

Fig. 1 shows the common approach to model contacts attached to the device.<sup>2,8</sup> The contact is treated as a semi-infinite quantum wire of finite cross-section where  $H_B$  and  $W$  represent the unit block Hamiltonian and coupling matrix between nearest blocks, respectively. Then  $\Psi_{(n,m)}$ , the eigenfunction in the  $n^{\text{th}}$  unit block of the plane wave at the  $m^{\text{th}}$  mode, will obey the Schrödinger equation and Bloch condition in (5), where  $k_m$  is the wave number of a plane-wave at the  $m^{\text{th}}$  mode and  $L$  is the size of one unit block along the direction of transport. In total, there are  $M$  modes where  $M$  is a dimension of  $H_B$ .

$$(EI - H_B)\Psi_{(n,m)} = W^+\Psi_{(n-1,m)} + W\Psi_{(n+1,m)}, \quad (5)$$

$$\Psi_{n+1} = \exp(ik_m L)\Psi_n, \quad 1 \leq m \leq M$$

Now one can solve the surface Green's function  $G_{surf}$  and self-energy  $\Sigma$  for the contact by formulating the general non-Hermitian eigenvalue problem with (5), where their general solutions are introduced in (6), (7).<sup>9</sup>

$$G_{surf} = K[K^{-1}(H_B - EI)K + K^{-1}W^+K\Lambda]^{-1}K^{-1}, \quad (6)$$

$$\Sigma = W^+G_{surf}W$$

$$K = [\Psi_{(0,1)} \quad \Psi_{(0,2)} \quad \dots \quad \Psi_{(0,M)}], \quad (7)$$

$$\Lambda = \text{diag}[\exp(-ik_1L), \dots, \exp(-ik_ML)]$$

The original CBR method prescribes  $X$  for  $\Sigma^x$  so that ( $H^o+X$ ) corresponds to the device Hamiltonian with Von Neumann boundary conditions.<sup>2</sup> This is feasible with the EM or  $k \cdot p$ <sup>12</sup> band model with cubic-grid bases, where the general expressions in (6) can be simplified to (8).

$$G_{surf} = -K\Lambda W^{-1}K^{-1}, \quad \Sigma = -WK\Lambda K^{-1} \quad (8)$$

However, the simplification in (8) is invalid with TB models, which use a set of atomistic bases with ZB grids. For further discussions, a simple example is used, where we assumed two [100] Si unit blocks with 1.2nm square cross-section. Fig. 2 shows its conceptual schematic and corresponding device Hamiltonian represented with the EM model (Top), and the spds\* TB approach (Bottom).

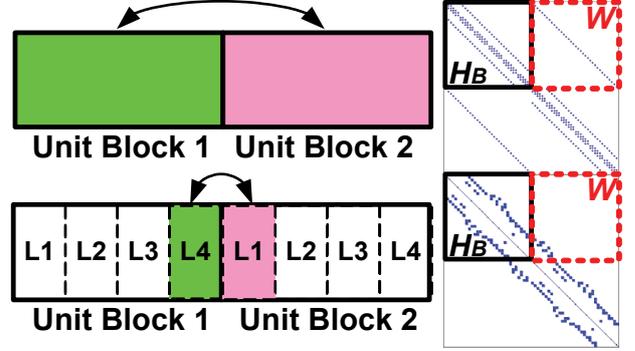


Figure 2. Conceptual illustration of the example structure and its Hamiltonian with EM (T) and TB Model (B).

In the EM model, each unit block has a common grid layer such that each layer always has non-zero couplings with next nearest one. However, a unit block of [100] Si has 4 unique atomic layers therefore, only the last layer in one unit block have non-zero couplings with the first layer in the nearest unit block, causing  $W$  to be a singular matrix as shown in Fig. 2. Then the simplification of (6) to (8) becomes difficult, since for square matrices  $K$  and  $W$ ,  $K^{-1}WK$  cannot be reduced to  $W$  and if  $W$  is singular.

Therefore, one needs to evaluate the self-energy term  $\Sigma$  with the solution in (6) with alternatives for  $X$  to make the CBR method still practical with TB models, one of which we suggested in (9), where  $\varepsilon_n$  is an eigenvalue at the  $n^{\text{th}}$  sub-band minima in the conduction band (maxima for the valence band) of the semi-infinite contact.

$$X = \frac{\Sigma(E = \varepsilon_n) + \Sigma(E = \varepsilon_n)^+}{2}, \quad \text{for } n^{\text{th}} \text{ sub-band} \quad (9)$$

If we are interested in the carrier transport through the first few contact sub-bands, which is the case to simulate RTD's or low-bias behaviors of quantum wires, the idea works very well because the  $X$  given in (9) represents the energy-independent term of the self-energy matrix at the specific sub-band, with which  $\Sigma^x$  becomes negligible at the vicinity of the corresponding sub-band minima.

## 2.2 Further optimization: Cases with two contacts

For the evaluation of  $G^R$ , CBR needs to calculate the inverse of the boundary block  $A_c$  in (3) which assumes the contact is coupled with the entire part of each unit

block in the boundary region of devices. In the ZB crystal structures, however, each unit block consists of a couple of unique and explicit atomic layers such that the requirement of matrix inversion can be further simplified. To measure the numerical efficiency of the suggested method with the RGF, we assume two contacts for the carrier transport in open system, which is the usual case in the modeling of quantum transport.<sup>5,9</sup>

As a detailed example, we assume a [100] Si quantum wire with 2 contacts in Fig. 3 and divide the device space into boundary region  $c=c_1+c_2$  and region  $d$  for the rest, where  $\Sigma_c^x$  in (3) is written as the expression in (10) since the unit block in the boundary region  $c_1$  is not coupled with the one in the region  $c_2$ . Then one can use (11) to calculate the transmission function  $T(E)$  with  $G_c^R$ , which is the boundary block of  $G^R$  and requires the inversion of a matrix of size  $2N$  for its evaluation where  $N$  represents the number of grid points in one unit block. The size of matrix inversions, however, can be further reduced with the idea suggested in Fig. 4.

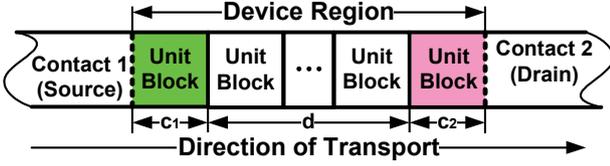


Figure 3. Schematic picture of the example device.

$$\Sigma_c^x = \begin{bmatrix} \Sigma_{c_1}^x & 0_{c_1 c_2}^x \\ 0_{c_2 c_1}^x & \Sigma_{c_2}^x \end{bmatrix} = \Sigma_S^x + \Sigma_D^x, \quad (10)$$

$$\Sigma_S^x = \begin{bmatrix} \Sigma_{c_1}^x & 0_{c_1 c_2}^x \\ 0_{c_2 c_1}^x & 0_{c_2}^x \end{bmatrix}, \quad \Sigma_D^x = \begin{bmatrix} 0_{c_1}^x & 0_{c_1 c_2}^x \\ 0_{c_2 c_1}^x & \Sigma_{c_2}^x \end{bmatrix}$$

$$T(E) = \text{trace}(\Gamma_S G_c^R \Gamma_D G_c^{R+}), \quad (11)$$

$$\Gamma_S = i(\Sigma_S^x - \Sigma_S^{x+}), \quad \Gamma_D = i(\Sigma_D^x - \Sigma_D^{x+})$$

$$T(E) = \text{trace}[\Sigma_{c(S)}^x G_{c(SD)}^R \Sigma_{c(S)}^R (G_{c(SD)}^R)^+], \quad (12)$$

$$G_{c(SD)}^R = (B_1 - B_2 B_3^{-1} B_4)^{-1} (G_{c(S)}^x - B_2 B_3^{-1} G_{c(D)}^x)$$

$$\text{where } B_1 = I - G_{c(S)}^x \Sigma_{c(S)}^x, \quad B_2 = -G_{c(SD)}^x \Sigma_{c(D)}^x,$$

$$B_3 = -G_{c(DS)}^x \Sigma_{c(S)}^x, \quad B_4 = I - G_{c(S)}^x \Sigma_{c(S)}^x$$

Here we decompose the boundary block  $\Sigma_c^x$ ,  $G_c^x$  with respect to four atomic layers in one [100] Si unit block. Then only (S) and (D) block of  $\Sigma_c^x$  will become non-zero because only the first layer of unit block  $c_1$  and the last layer of unit block  $c_2$  will be coupled with the source and drain contact, respectively. Then, one can easily convert the equations for  $T(E)$  in (11) to (12) with simple matrix

arithmetic, and the evaluation of  $T(E)$  can be done with matrix inversion of size  $N/4$ , which corresponds to the number of grid points in only one of the 4 atomic layers of each [100] Si unit block. We note the size for matrix inversion now becomes 8 times smaller than the previous requirement ( $2N$ ), which results in a huge reduction of computations, as will be discussed in the next section.

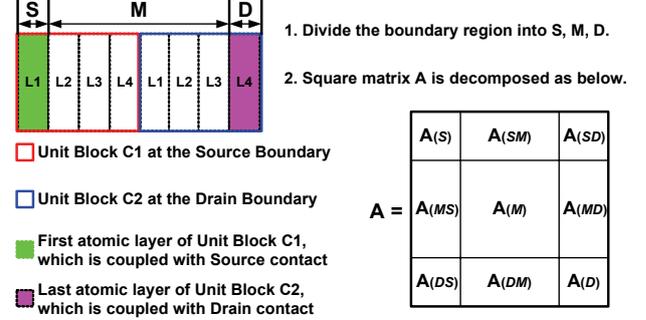


Figure 4. A rule to decompose boundary block matrices.

### 3. Result

In this section, we show the computational practicality of the methodology discussed so far, by investigating the tunneling behavior of electrons through a single impurity atom placed in the channel of a Si quantum wire, which is important in the tunneling spectroscopy to understand the electronic structure of low-dimensional systems.<sup>13</sup>

#### 3.1 Description of numerical example

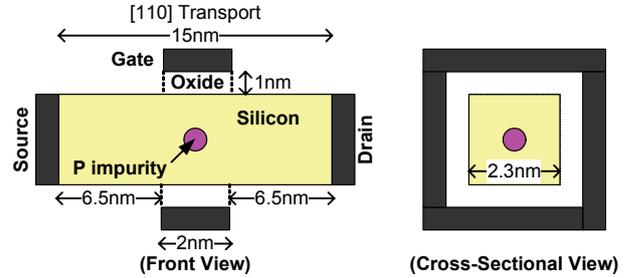


Figure 5. Schematic of the device for simulation.

A schematic of the target device is described in Fig. 5. Here the wire channel has a length of 15nm with a 2.3nm rectangular cross-section. For [110] transport, we assume the source and drain contact with a line-doping constant of  $10^8$  (donors/m) where the gate contact with 1nm oxide layer was used to consider the band bending along the cross-sectional direction. Then a single phosphorous ion is placed in the channel with an analytical consideration of the impurity potential.<sup>14</sup> This numerical example uses the semi-empirical  $sp^3d^5s^*$  TB model,<sup>15</sup> and the size of the corresponding device Hamiltonian is 23,010.

### 3.2 Computational efficiency of the methodology

For a measurement of the numerical efficiency, the device is simulated for 4 different cases as shown in table 1, where the RGF method was used as a reference. Here we only considered the transport of electrons in the conduction band such that the transmission function was evaluated with the first few conduction band eigenstates of  $(H^o+X)$ . The test has been performed with MATLAB codes on a Power Mac G5 consisting of dual 1.8 GHz CPUs and 2GB of SDRAM.

Table 1. Four different approaches for the simulation

Method	Number of used ev's	Subdivision of boundary block	Time for sim. (s)
A: CBR	3(0.013%)	Use	658
B: CBR	10(0.044%)	Use	647
C: CBR	10(0.044%)	Don't Use	3341
D: RGF	----	----	7737

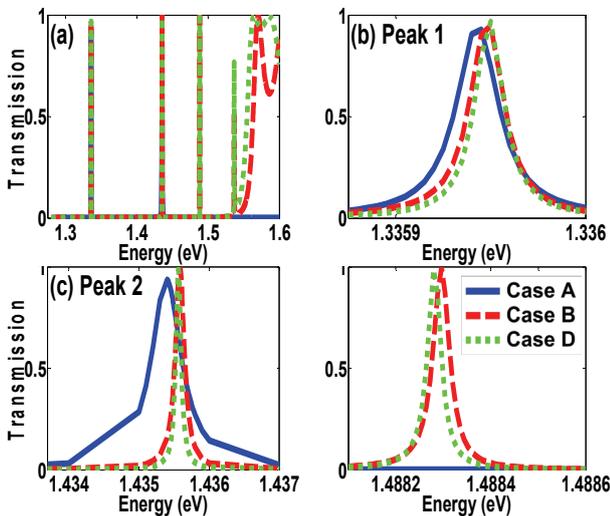


Figure 6. Transmission profiles. (a) entire energy range (b) near the 1<sup>st</sup>, (c) 2<sup>nd</sup>, and (d) 3<sup>rd</sup> resonance peak.

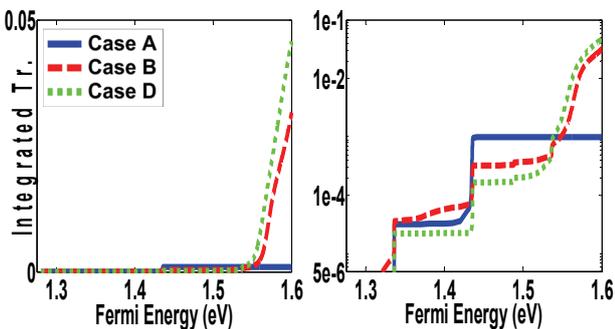


Figure 7. Integrated transmission profiles with respect to the contact fermi-level represented with a linear (Left), and a log scale (Right).

Fig. 6 shows the transmission profiles for Cases A, B, and D, with the close-up results focusing on the first 3 resonance peaks, where the CBR method gives a result closer to the reference if one uses more eigenstates. The case B with 10 eigenstates reproduces the reference over almost the entire energy range of our interest. The case A still shows a good performance at the vicinity of the first resonance, but the result starts to deviate near the second peak, and finally vanishes. The accuracy of results is also supported with Fig. 7, where we integrated transmission functions with respect to the fermi-energy of contacts in equilibrium. The required times for different simulations, the details of which were summarized in table 1, indicate that the alleviation of matrix inversions indeed results in a huge reduction of computing expenses.

### 4. Summary

The CBR method is shown to be practical to compute the retarded Greens' function for open devices with atomistic band models. A 3D nanowire with an embedded impurity which exhibits tunneling behaviors with extremely sharp resonances, is used to demonstrate that the transmission can be computed with few eigenstates of the closed system. The matrix inversion needed to compute the transmission can be reduced such that one can save significant computing costs.

### Acknowledgments

We acknowledge support from the National Science Foundation under contract No. 0701612 and the Semiconductor Research Corporation. nanoHUB. computational resources have been used in this work.

### References

- [1] D. Mamaluy *et al.*, *Phys. Rev. B* **71**, 245321 (2005).
- [2] D. Mamaluy *et al.*, *Semicond. Sci. Tech.* **19**, 118 (2004).
- [3] H. Khan *et al.*, *J. Comput. Electron* **6**, 77 (2007).
- [4] M. Luisier *et al.*, *Phys. Rev. B* **74**, 205323 (2006).
- [5] G. Klimeck *et al.*, *IEEE Trans. Electron Devices* **54**, 2090 (2007).
- [6] G. Klimeck *et al.*, *VLSI Design* **8**, 79 (1998).
- [7] P. Vogl *et al.*, *J. Phys. Chem. Solids* **44**, 365 (1983).
- [8] S. Datta, *Superlatt and Microstruct.* **28**, 253 (2000).
- [9] C. Rivas *et al.*, *Phys. Stat. Sol. (b)* **239**, 94 (2003).
- [10] R. Lake *et al.*, *J. Appl. Phys.* **81**, 7845 (1996).
- [11] S. Cauley *et al.*, *J. Appl. Phys.* **101**, 123715 (2007).
- [12] Y. X. Liu *et al.*, *Phys. Rev. B* **54**, 5675 (1996).
- [13] E. Lind *et al.*, *Phys. Rev. B* **68**, 033312 (2003).
- [14] R. Rahman *et al.*, *Phys. Rev. Lett.* **99**, 036403 (2007).
- [15] T. B. Boykin *et al.*, *Phys. Rev. B* **71**, 115201 (2004).