The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	**Efficient simulation and analysis of quantum ballistic transport in nanodevices with AWE**
Author(s)	**Huang, JZ; Chew, WC; Tang, M; Jiang, L**
Citation	**Ieee Transactions On Electron Devices, 2012, v. 59 n. 2, p. 468-476**
Issued Date	**2012**
URL	**http://hdl.handle.net/10722/155724**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

Efficient Simulation and Analysis of Quantum Ballistic Transport in Nanodevices With AWE

Jun Z. Huang, Weng Cho Chew, *Fellow, IEEE*, Min Tang, *Member, IEEE*, and Lijun Jiang, *Member, IEEE*

Abstract—Quantum–mechanical modeling of ballistic transport in nanodevices usually requires solving the Schrödinger equation at multiple energy points within an energy band. To speed up the simulation and analysis, the asymptotic waveform evaluation is introduced in this paper. Using this method, the wave function is only rigorously solved at several sampled energy points, whereas those at other energies are computed through Padé approximation. This allows us to obtain the physical quantities over the whole energy band with very little computational cost. In addition, the accuracy is controllable by a complex frequency hopping algorithm. The validity and efficiency of the proposed method are demonstrated by detailed study of several multigate silicon nano-MOSFETs.

Index Terms—Asymptotic waveform evaluation (AWE), complex frequency hopping (CFH), local density of states (LDOS), multigate MOSFET, nanodevices, quantum transport, Schrödinger equation.

I. INTRODUCTION

SINCE the dimensions of nanodevices have shrunk to be comparable to electron wavelength, quantum–mechanical modeling of electron transport through these nanodevices is indispensable to capture their wave-physics features. Several quantum transport models have been developed with different levels of approximations [1]. To calculate ballistic current through ultrasmall nanostructures, a widely used scheme is to solve the coupled Schrödinger–Poisson system self-consistently, either directly [2]–[4] or by using a nonequilibrium Green's function (NEGF) approach [5]–[7].

Both methods essentially generate the same results. In terms of computational burden, the former approach usually requires less computer time than the latter one, because the wave function is directly computed for each mode coming from the contact and the number of modes with energy below Fermi level is usually very small. However, the NEGF approach is quite

convenient since all the modes in the contacts are automatically taken into account in the Green's function and all the physical quantities are expressed in very compact forms.

Solving the self-consistent Schrödinger–Poisson system is computationally intensive, as it requires solving the open boundary Schrödinger equation for each energy point, each incoming mode, each iteration, and each bias. There are several efficient models developed over the past years to reduce the complexity of the large matrix inversion, including the quantum transmitting boundary methods [2], the coupled (and uncoupled) mode space approach [3], [5], the scattering matrix method [8], the recursive Green's function method [9], the contact block reduction method [10], [11], and the R-matrix method [12]–[14]. However, solving the Schrödinger equation repeatedly at every energy point is still time consuming, and there is a need to find approximate solutions that can efficiently simulate the energy response over a wideband.

Asymptotic waveform evaluation (AWE) combined with the CFH technique is a very popular frequency sweep method in high speed circuit analysis and computational electromagnetics [15], which has been verified to be able to reduce the computer time by over one order of magnitude. In this paper, this technique is employed for efficient simulation and analysis of quantum electron transport in nanodevices. It will be shown that this method considerably speeds up the simulation while good accuracy can be maintained. In Section II, quantum ballistic transport equations will be reviewed first, and then the idea of AWE and CFH will be presented and incorporated into the simulation flow. In Section III, the new method is applied to simulate and analyze several multigate silicon MOSFETs; the accuracy and simulation time are compared with traditional approach. Some conclusions are drawn in Section IV.

II. METHOD DESCRIPTION

A. Quantum Ballistic Transport Problem

A general 2-D quantum device is illustrated in Fig. 1. The solution domain we are interested in can be divided into two parts, namely, the device region Ω_D and the contact regions Ω_α ($\alpha = 1, 2,$ and 3). Ω_D is a finite region with a boundary denoted by Γ_D; Ω_α is a semi-infinite region with a boundary denoted by Γ_α. Denote the intersection of Ω_D and Ω_α by $\Gamma_{D,\alpha}$, then the rest of Γ_D is $\Gamma_{D,0}$, and the rest of Γ_α is $\Gamma_{\alpha,0}$.

The device region is characterized by space varying potential $V_D(x, y)$ and effective mass $m_D^*(x, y)$. Since the potential and effective mass inside the contact region should be independent of the position along the contact although it may have a complicated transverse structure [2], the potential
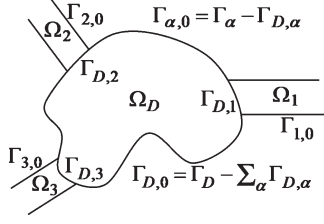
Fig. 1. Geometry of a generalized 2-D quantum device with three semi-infinite leads.

$V_\alpha(\xi_\alpha, \eta_\alpha) = V_\alpha(\xi_\alpha)$ and $m_\alpha^*(\xi_\alpha, \eta_\alpha) = m_\alpha^*(\xi_\alpha)$, where ξ_α and η_α are respectively the transverse position and the longitude position inside each contact. Therefore, our major problem is [2]:

Given: (1) the potential and effective mass in every region: $V_D(x, y)$, $m_D^*(x, y)$, $V_\alpha(\xi_\alpha)$, and $m_\alpha^*(\xi_\alpha)$ and (2) amplitude for each wave incoming from contact α with mode n and energy E,

Find: $\psi_D^{\alpha,n}(x, y, E) \in C^2(\Omega_D)$, which satisfies the following stationary Schrödinger equation:

$$-\frac{\hbar^2}{2} \nabla \cdot \left[\frac{1}{m_D^*(x, y)} \nabla \right] \psi_D^{\alpha,n}(x, y)$$

$$+ V_D(x, y) \psi_D^{\alpha,n}(x, y) = E \psi_D^{\alpha,n}(x, y), (x, y) \in \Omega_D \quad (1)$$

and boundary conditions:

$$\psi_D^{\alpha,n} = \psi_\alpha \text{ on } \Gamma_{D,\alpha} \quad (2)$$
$$\nabla \psi_D^{\alpha,n} \cdot \vec{n}_{\Gamma_{D,\alpha}} = \nabla \psi_\alpha \cdot \vec{n}_{\Gamma_{D,\alpha}} \text{ on } \Gamma_{D,\alpha} \quad (3)$$
$$\psi_D^{\alpha,n} = 0 \text{ on } \Gamma_{D,0} \quad (4)$$
$$\psi_\alpha = 0 \text{ on } \Gamma_{\alpha,0} \quad (5)$$
$$\psi_\alpha \text{ bounded as } \eta_\alpha \to \infty. \quad (6)$$

Once the wave function is obtained by solving (1)–(6), all the physical quantities can be obtained. For example, the electron density is given by

$$n(x, y) = 2 \sum_\alpha \sum_n \int_0^{+\infty} |\psi_D^{\alpha,n}(x, y, E)|^2$$

$$\times f_{\text{FD}}(E - \mu^\alpha) \frac{dk}{dE} \frac{dE}{2\pi} \quad (7)$$

where k is the wavenumber, f_{FD} is the Fermi–Dirac distribution function, and μ_α is the Fermi level associated to contact α. Note that incoming wave of different modes or contacts are uncorrelated, so they are independently calculated and added up [16]. The terminal current can be obtained through transmission function by the Landauer–Büttiker formula

$$I_\alpha = \frac{2q}{h} \sum_{\alpha \neq \alpha'} \int_0^{+\infty} T_{\alpha\alpha'}(E)$$

$$\times \left[f_{\text{FD}}(E - \mu^\alpha) - f_{\text{FD}}(E - \mu^{\alpha'}) \right] dE \quad (8)$$

$$T_{\alpha\alpha'}(E) = \sum_n \sum_m \frac{k_m^\alpha}{k_n^{\alpha'}} \left| \psi_D^{\alpha',n}(x, y, E)^\dagger \cdot \chi_m^\alpha(\xi_\alpha) \right|^2 \quad (9)$$

where $\chi_m^\alpha(\xi_\alpha)$ is the mth normalized eigenmode of the contact α, which will be defined later on.

It should be mentioned that the potential distribution for (1) is usually determined by a self-consistent procedure, which requires solving the Poisson equation with the charge density obtained from (7), i.e.,

$$\nabla \cdot [\epsilon(x, y) \nabla V_D(x, y)] = q [n(x, y) - N_d(x, y)] \quad (10)$$

where ϵ is the dielectric constant, N_d is the doping density, and q is the electron charge. The boundary conditions for the Poisson equation will be specified later for the specific device.

B. Numerical Solution

To obtain the wave function in the device region, we first write down the solution in the contact regions for one incoming mode as the summation of incident and scattered waves. Thus

$$\psi_\alpha(\xi_\alpha, \eta_\alpha) = a_n^\alpha \chi_n^\alpha(\xi_\alpha) \exp(-i k_n^\alpha \eta_\alpha)$$

$$+ \sum_{m=1}^{N\alpha} b_m^\alpha \chi_m^\alpha(\xi_\alpha) \exp(i k_m^\alpha \eta_\alpha) \quad (11)$$

where a_n^α and b_m^α are the amplitudes of incident wave and scattered wave, respectively. Here, $\chi_m^\alpha(\xi_\alpha)$ is the mth normalized eigenmode of the contact that satisfies the following eigenvalue problem (suppose $m_\alpha^*(\xi_\alpha)$ is constant in the contacts) and boundary condition (5):

$$-\frac{\hbar^2}{2m_\alpha^*} \frac{\partial^2}{\partial \xi_\alpha^2} \chi_m^\alpha(\xi_\alpha) + V_\alpha(\xi_\alpha) \chi_m^\alpha(\xi_\alpha) = E_m^\alpha \chi_m^\alpha(\xi_\alpha) \quad (12)$$

which can be numerically solved. Here, k_m^α is the longitudinal wavenumber, and

$$k_m^\alpha = \sqrt{2m^* (E - E_m^\alpha)}/\hbar. \quad (13)$$

It should be noted that k_m^α can be either real or imaginary, which corresponds to travelling wave or evanescent wave in the contact. Here, N_α should be truncated to include enough number of evanescent waves.

According to the orthogonality of the eigenmodes, we can evaluate b_m^α as

$$b_m^\alpha = \int \chi_m^\alpha(\xi_\alpha) \psi_\alpha(\xi_\alpha, \eta_\alpha = 0) d\xi_\alpha - a_n^\alpha \delta_{mn}. \quad (14)$$

Substituting the above expression back to (11) and using boundary condition (2), we have

$$\psi_\alpha(\xi_\alpha, \eta_\alpha) = -2i a_n^\alpha \chi_n^\alpha(\xi_\alpha) \sin(k_n^\alpha \eta_\alpha)$$

$$+ \sum_{m=1}^{N\alpha} \left(\int \chi_m^\alpha(\xi_\alpha) \psi_D^{\alpha,n}(\xi_\alpha, 0) d\xi_\alpha \right) \chi_m^\alpha(\xi_\alpha) \exp(i k_m^\alpha \eta_\alpha).$$

$$(15)$$

This is the solution in the contact region in terms of the unknowns at the interface; it is subsequently utilized to express the boundary conditions for the device region.

Next, applying the finite-difference method to discretize the 2-D Schrödinger equation (1), we can obtain the following matrix equation:

$$\left[E\mathbf{I} - \mathbf{H} - \sum_{\alpha} \mathbf{S}^{\alpha}(E) \right] \mathbf{\Psi}_D^{\alpha,n}(E) = \mathbf{v}_n^{\alpha}(E) \qquad (16)$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{H}$ is the isolated device Hamiltonian matrix, $\mathbf{S}^{\alpha}(E)$ is the energy dependent self energy matrix that represents the boundary condition, and $\mathbf{v}_n^{\alpha}(E)$ is the vector that represents the incident wave from the contact. $\mathbf{S}^{\alpha}(E)$ and $\mathbf{v}_n^{\alpha}(E)$ only have nonzero elements in the parts that have coupling to the contacts, which can be derived from (15) and are given in Appendix I.

Equation (16) can then be solved by various matrix solvers. It must be pointed out that, for multiple incoming modes from multiple contacts that share the same energy, they have the same Hamiltonian matrix that only needs to be inverted once.

C. AWE

It is obvious from (7) and (8) that (16) needs to be repeatedly solved within the energy range of interest so as to obtain the integral. In particular, when the wave function rapidly changes with energy, the energy grid must be very fine so as to achieve convergence. In addition, sometimes we need to plot the spectrum of the electron density or the transmission coefficients over the energy range to analyze the device physics and guide the design process. These can be very time consuming for large problems.

To obtain the solution of (16) over a wide energy band, following the steps in [15], we rewrite (16) as

$$\mathbf{A}(E)\mathbf{\Psi}(E) = \mathbf{v}(E) \qquad (17)$$

and expand $\mathbf{\Psi}(E)$ in terms of Taylor series at E_0, i.e.,

$$\mathbf{\Psi}(E) \approx \sum_{n=0}^{Q} \mathbf{m}_n (E - E_0)^n. \qquad (18)$$

Similarly, we expand $\mathbf{A}(E)$ and $\mathbf{v}(E)$ in terms of Taylor series at E_0 with coefficients $\mathbf{A}^{(n)}$ and $\mathbf{v}^{(n)}$. Matching the coefficients of equal powers on both sides of (17) leads to the following recursive algorithm for $\mathbf{m}_n$:

$$\mathbf{m}_0 = \mathbf{A}^{-1}(E_0)\mathbf{v}(E_0) \qquad (19)$$

$$\mathbf{m}_n = \mathbf{A}^{-1}(E_0) \left[\frac{\mathbf{v}^{(n)}(E_0)}{n!} - \sum_{i=1}^{n} \frac{\mathbf{A}^{(i)}(E_0)\mathbf{m}_{n-i}}{i!} \right], \quad n \geq 1. \qquad (20)$$

A wider bandwidth can be obtained by approximate $\mathbf{\Psi}(E)$ with a rational Padé approximant of order $[L/M]$, i.e.,

$$\mathbf{\Psi}(E) \approx \frac{\sum_{i=0}^{L} \mathbf{a}_i (E - E_0)^i}{1 + \sum_{j=1}^{M} \mathbf{b}_j (E - E_0)^j} \qquad (21)$$

where $L + M = Q$. The elements of unknown coefficient vectors $\mathbf{a}_i (0 \leq i \leq L)$ and $\mathbf{b}_j (1 \leq j \leq M)$ can be calculated by

equating the right-hand sides of (18) and (21), multiplying both sides with the denominator of the Padé approximant, and matching the coefficients of equal powers of $E - E_0$, which results in

$$\begin{bmatrix} m_L & m_{L-1} & m_{L-2} & \cdots & m_{L-M+1} \\ m_{L+1} & m_L & m_{L-1} & \cdots & m_{L-M+2} \\ m_{L+2} & m_{L+1} & m_L & \cdots & m_{L-M+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{L+M-1} & m_{L+M-2} & m_{L+M-3} & \cdots & m_L \end{bmatrix}$$
$$\bullet \begin{bmatrix} b_1 & b_2 & b_3 & \cdots & b_M \end{bmatrix}^T$$
$$= -\begin{bmatrix} m_{L+1} & m_{L+2} & m_{L+3} & \cdots & m_{L+M} \end{bmatrix}^T \qquad (22)$$

and

$$a_i = \sum_{j=0}^{i} b_j m_{i-j}, \qquad 0 \leq i \leq L. \qquad (23)$$

Equation (22) is first solved to obtain b_j, which is then substituted into (23) to calculate a_i.

Once coefficient vectors $\mathbf{a}_i$ and $\mathbf{b}_j$ are evaluated, the wave function at any energy (within the bandwidth of accuracy) can be found by (21). Note that if we do LU decomposition of sparse matrix $\mathbf{A}$, then (19) and (20) can be efficiently solved with forward and backward substitutions. The implementation is simple, since the derivatives of $\mathbf{A}$ only have nonzero elements in the self-energy parts, and they have very simple analytical forms as can be derived from (A.5) and (13). Similarly, the derivatives of $\mathbf{v}$ only have nonzero elements in the layer that couples to the contact, and they are also analytical.

In addition, for multiple incoming modes from multiple contacts, we need to evaluate the asymptotic form (21) for each mode. Note that the LU decomposition of $\mathbf{A}$ can be reused, and the computational cost is slightly increased since we need to do more forward and backwards substitutions.

D. Complex Frequency (Energy) Hopping

Since the bandwidth of Padé approximation is limited, multiple points' expansion is necessary to obtain an accurate solution over the whole energy band. The locations of the energy points can be selected by the CFH technique, as described in the following.

Given an energy band $[E_1, \mu + 4k_B T]$, where E_1 is the eigen energy of the incoming wave, μ is the Fermi level, k_B is the Boltzmann constant, and T is the temperature in Kelvin, and a maximum error tolerance ε for the wave function $\mathbf{\Psi}$, the following process is observed.

1) Let $E_{\min} = E_1$ and $E_{\max} = \mu + 4k_B T$.
2) Do AWE at $E_{\min}$ and $E_{\max}$, obtain $\mathbf{\Psi}_1(E)$ and $\mathbf{\Psi}_2(E)$.
3) Calculate $\mathbf{\Psi}_1(E_{\mathrm{mid}})$ and $\mathbf{\Psi}_2(E_{\mathrm{mid}})$, respectively, at middle energy $E_{\mathrm{mid}} = (E_{\min} + E_{\max})/2$.
4) If $\max|\mathbf{\Psi}_1(E_{\mathrm{mid}}) - \mathbf{\Psi}_2(E_{\mathrm{mid}})| < \varepsilon$, stop. Otherwise, do AWE at E_{mid} and repeat the above steps for subregions $[E_{\min}, E_{\mathrm{mid}}]$ and $[E_{\mathrm{mid}}, E_{\max}]$.

However, a numerical difficulty arises in the above process because the derivatives reach singularities at subband edges
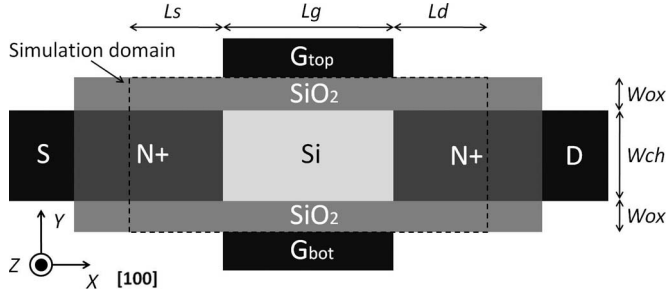
Fig. 2. Two-dimensional view of the n-type double-gate silicon MOSFET. The structure is infinite in the Z direction. Gate length is denoted by L_g; source and drain extension lengths are denoted by L_s and L_d, respectively; silicon channel thickness is W_{ch}; and oxide thickness is W_{ox}.
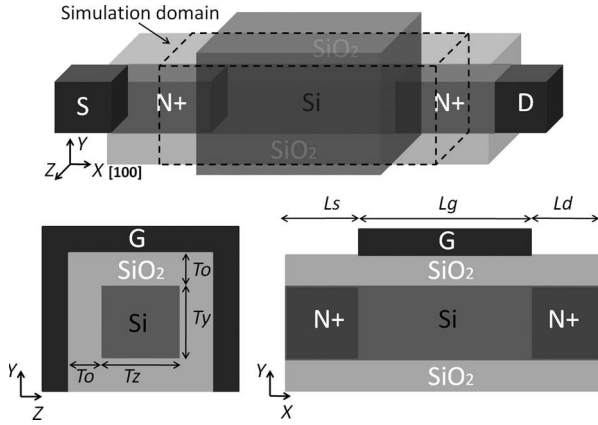

Fig. 3. (Top) Three-dimensional view, (bottom left) YZ cross section view, and (bottom right) XY cross section view of the n-type triple-gate silicon MOSFET. Gate length is denoted by L_g; source and drain extension lengths are denoted by L_s and L_d, respectively; silicon channel thickness is T_x and T_y; and oxide thickness is T_o.

$E = E_m^\alpha$, as shown in (13). In order to avoid the singularities, several small intervals should be skipped from the expansion region.

Suppose there are N subbands within region $[E_1, \mu + 4k_BT]$, the subband edge energies of which are

$$E_1 < E_2 < \cdots < E_N < \mu + 4k_BT. \quad (24)$$

Then, we divide the whole region into several subregions

$$(E_1 + \sigma, E_2 - \sigma), (E_2 + \sigma, E_3 - \sigma), \ldots, (E_N + \sigma, \mu + 4k_BT) \quad (25)$$

where σ is a small value, e.g., 5×10^{-4} eV. In each subregion, the CFH algorithm is employed as described above.

III. NUMERICAL RESULTS AND DISCUSSION

The above proposed method is applied to simulate several 2-D and 3-D multigate silicon MOSFETs, as shown in Figs. 2 and 3. They are very promising candidates for the next generation nanotransistors.

For the Poisson equation, the Dirichlet boundary condition is enforced at the gate region, whereas the floating boundary condition, i.e., the normal derivative is zero, is applied at the remaining boundary. This can be used to maintain the charge neutrality at the source and drain extensions [17]. In addition,
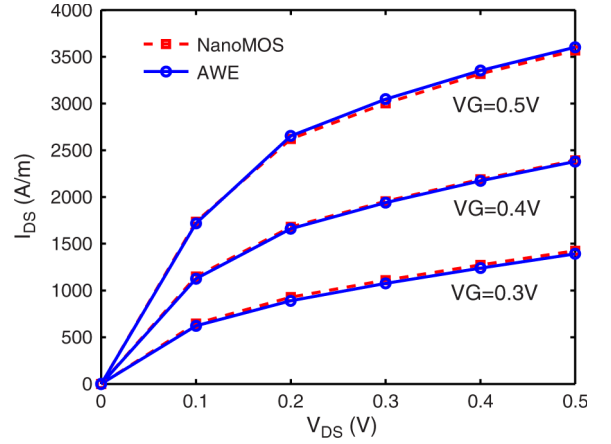

Fig. 4. Currents of the double-gated MOSFET as a function of drain bias for different gate voltages ($V_G = 0.3, 0.4, 0.5$ V): comparison between results calculated by AWE and the results of nano-MOS.
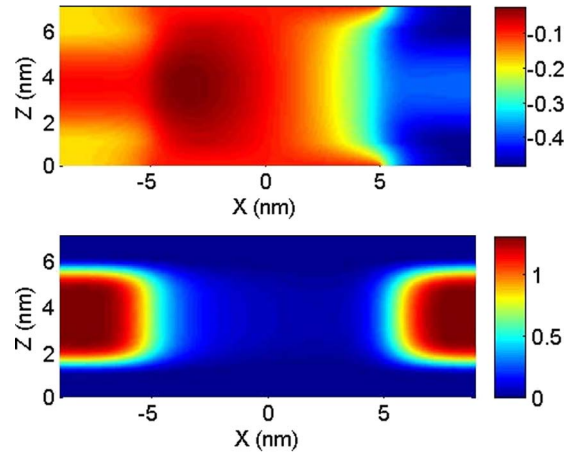

Fig. 5. Two-dimensional plot of the (top) potential and (bottom) electron density distributions at $V_G = V_D = 0.3$ V.

the Gummel iterative scheme is adopted to speed up the convergence of the coupled Schrödinger–Poisson system [3].

A. Two-Dimensional Double-Gate MOSFET

As shown in Fig. 2, the device parameters are $L_g = 10$ nm, $L_s = L_d = 4$ nm, $W_{ch} = 5$ nm, and $W_{ox} = 1$ nm; doping density is $N^+ = 10^{26}/m^3$; longitudinal and transverse effective mass are $m_l^* = 0.91 m_e$ and $m_t^* = 0.19 m_e$; work function of gate metal is 4.25 eV; affinity of silicon is 4.05 eV; permittivity of silicon is 11.9; and permittivity of SiO$_2$ is 3.8. Temperature $T = 300$ K. The grid spacing is 0.1 nm in both x and y directions. The source Fermi level is set to be 0 eV. In the following simulation, we use Padé approximant of order [4/4] and set the error tolerance of CFH to be $2 \times 10^{-2} \times \max|\chi_n^\alpha|$.

At first, we verified our code by comparing the I–V curve with that generated by the nano-MOS tool [18], which is a program using mode space NEGF formalism. It is shown in Fig. 4 that good agreement is obtained at every bias point.

The self-consistent potential and electron density distribution are plotted in Fig. 5. It is shown that the electron density is very high at the source and drain extensions, and thus, charge neutrality should be achieved because of the heavy positive
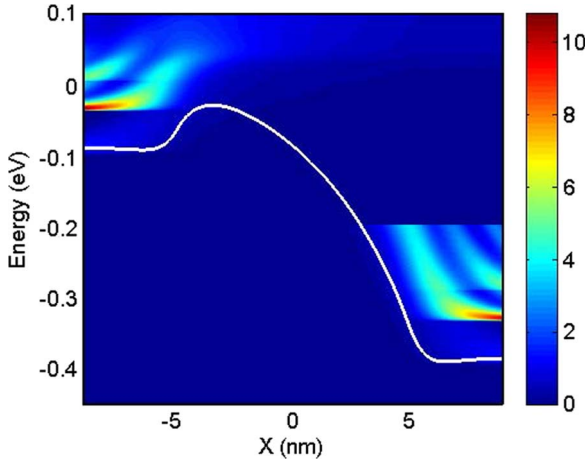
Fig. 6. LDOS along the center of the silicon layer. $V_G = V_D = 0.3$ V. Conduction band edge is also shown in white line.
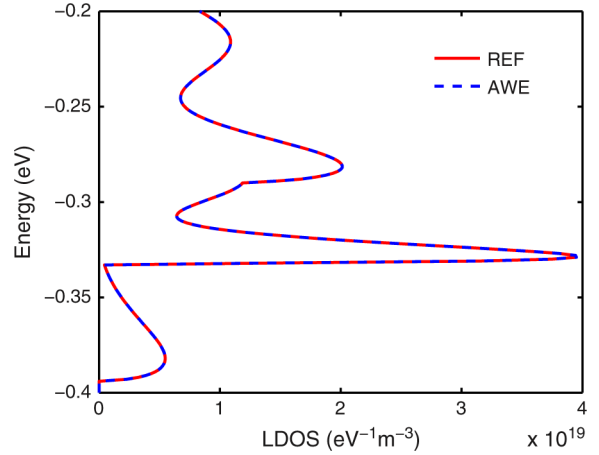

Fig. 8. LDOS at the center of the drain end; both reference and AWE results are plotted. $V_G = V_D = 0.3$ V.
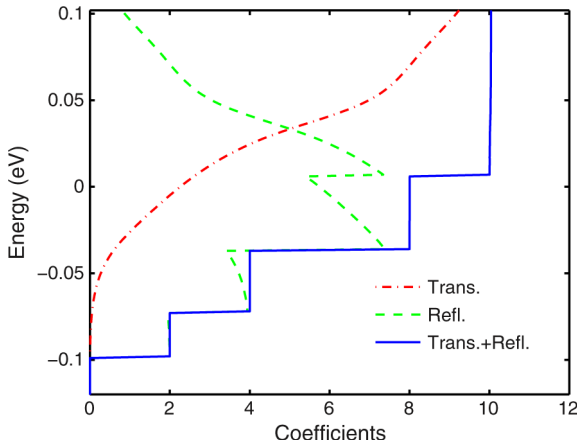

Fig. 7. Transmission and reflection coefficients defined in the Landauer formula for the electrons coming from the source. $V_G = V_D = 0.3$ V.
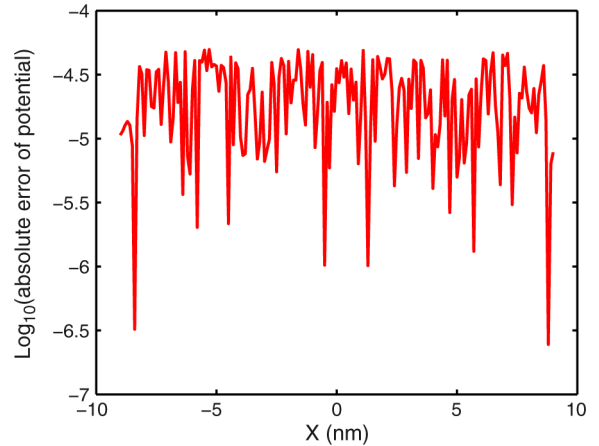

Fig. 9. Absolute error of the potential energy along the center of the silicon layer with $V_G = V_D = 0.3$ V.

doping density. The homogeneous Neumann boundary condition makes the potential constant along the transport direction at the source and drain ends.

The local density of states (LDOS) along the center of the silicon layer is depicted in Fig. 6. It is observed that the interference of incoming wave and reflected wave leads to a standing-wave-like phenomenon. In addition, some high energy electrons coming from the source side may go into the channel and eventually escape to the drain side, but the electrons coming from the drain can hardly go to the source side due to a large potential barrier. Therefore, current is formed with a direction from right to left.

Transmission and reflection coefficients are plotted in Fig. 7. We notice that the transmission coefficient is continuous, whereas the reflection coefficient jumps when a new mode starts to propagate. This is because when an electron mode starts to propagate, most of it will be reflected back. However, the summation of transmission and reflection coefficients is always equal to the number of propagating modes of that energy, which is an integer. The jumps of coefficient from 0 to 2, 2 to 4, and 8 to 10 correspond to the increase in propagating modes of the electrons with heavy effective mass in the confinement direction; their valley degeneracy is 2. Conversely, the jump of

coefficient from 4 to 8 corresponds to the propagating mode of electrons with light effective mass in the confinement direction; their valley degeneracy is 4.

To investigate the accuracy of the proposed method, we take the results of the direct method (calculating the values at each energy point) with a very fine energy grid (energy step: 0.001 eV) as the reference. Fig. 8 gives the LDOS in the middle of the drain end; it is shown that our method can produce almost the same results as the reference one. Fig. 9 plots the absolute error of the potential along the center of the silicon layer; it is shown that the error can be controlled below 10^{-4} V, which is very accurate.

We then investigate the performances of various orders of Padé approximant with different CFH tolerances, in comparison with the direct method (energy grid: 0.001 eV). The number of energy points (for AWE, it should be understood as the expansion points) of the last Poisson–Schrödinger iteration, total CPU time, and drain current for one bias point ($V_G = V_D = 0.3$ V) are summarized in Table I (matrix solver: sparse LU decomposition with permutation matrices P and Q using UMFPACK routines). Compared with the direct method, our method reduces the inversion points by over one order of magnitude. More accurate results can be obtained by reducing

	Energy Points	Current (A/m)	CPU time (seconds)	Speed Up
Direct method	951	1075.1	1413	$1.0\times$
$\varepsilon = 2\times10^{-2}$, Order [3/3]	73	1075.0	196	$7.2\times$
$\varepsilon = 2\times10^{-2}$, Order [4/4]	65	1075.1	200	$7.1\times$
$\varepsilon = 2\times10^{-2}$, Order [5/5]	55	1075.0	204	$6.9\times$
$\varepsilon = 4\times10^{-2}$, Order [3/3]	67	1074.9	185	$7.6\times$
$\varepsilon = 4\times10^{-2}$, Order [4/4]	62	1075.1	196	$7.2\times$
$\varepsilon = 4\times10^{-2}$, Order [5/5]	53	1075.0	203	$7.0\times$
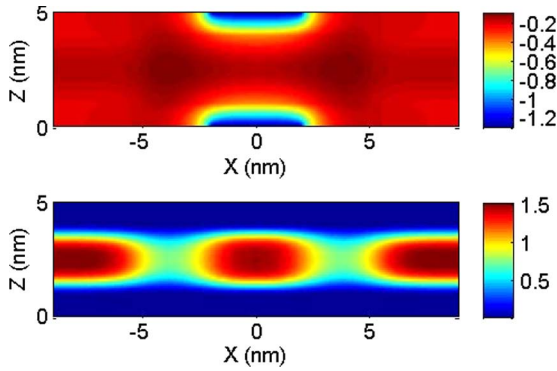
Fig. 10. Two-dimensional plot of the (top) potential and (bottom) electron density distributions at $V_G = 1.5$ V and $V_D = 0$ V.

the CFH tolerance but at the cost of more computer time since more expansion points are needed. To achieve the same accuracy, higher order Padé approximant takes more computational cost, although the number of expansion points decreases with increasing order. This is because higher order needs more forward and backward substitutions. When $\varepsilon = 4 \times 10^{-2}$, the drain currents obtained are still accurate enough, and it is over seven times faster.

B. Two-Dimensional Double-Gate (Underlapped) MOSFET

To further demonstrate the advantage of AWE, we analyze the same structure as Fig. 2, but this time, the gate length is reduced to 4 nm (note that the channel length is 10 nm), which means it is underlapped. The other device parameters are the same as those in Section A, except that $W_{\text{ch}} = 3$ nm.

The self-consistent potential and electron density distributions are plotted in Fig. 10. It is shown that the potential in the middle part of the channel is significantly lowered by the large gate bias, whereas the potential at the end parts of the channel is less affected by the gate bias, and correspondingly, the electron density only concentrates at the middle part of the channel.

We further plot the LDOS and conduction band edge along the center of the silicon layer in Fig. 11; it is obvious that the wave can penetrate through the potential barriers. In addition, because of the two potential barriers formed at the channel ends, there exist some resonant states inside the channel. The transmission and reflection coefficients are plotted in Fig. 12. It shows several sharp peaks corresponding to the resonant tunneling behavior. These sharp peaks are well captured by our
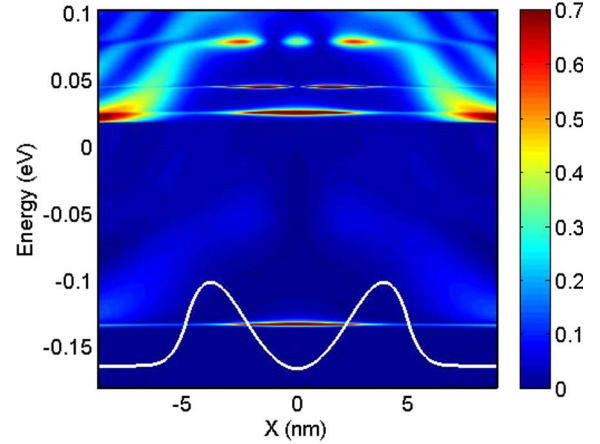
Fig. 11. LDOS along the center of the silicon layer. $V_G = 1.5$ V, $V_D = 0$ V. Conduction band edge is also shown in white line.
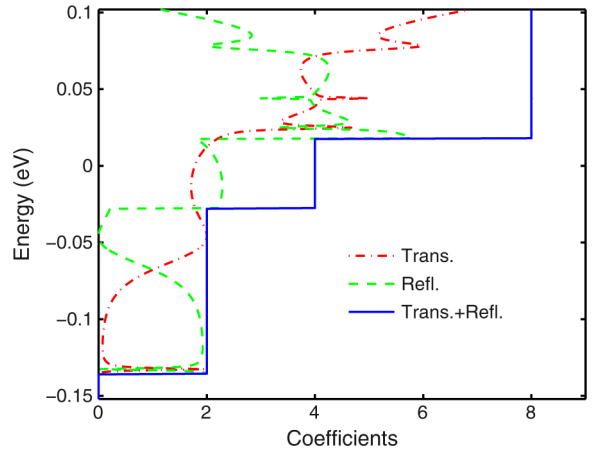
Fig. 12. Transmission and reflection coefficients defined in the Landauer formula for the electrons coming from the source. $V_G = 1.5$ V, $V_D = 0$ V.

method, which is hard to obtain by the direct method because it requires very fine energy grids.

C. Three-Dimensional Triple-Gate MOSFET

A triple-gate silicon MOSFET is simulated in this example, as shown in Fig. 3. The device parameters are $L_g = 10$ nm, $L_s = L_d = 4$ nm, $T_o = 1$ nm, and $T_y = T_z = 3$ nm. Due to the small cross section of the silicon nanowire, the effective masses are chosen as those in [19], which are extracted from $sp^3d^5s^*$ tight binding calculation of the $E - k$ dispersion. The other parameters are the same as those in Section A. The grid spacing is 0.2 nm in all x, y, and z directions.

The potential and electron density distributions are plotted in Figs. 13 and 14, respectively. It is shown that the potential in the XY plane is asymmetrical, whereas the potential in the XZ plane is symmetrical due to the tri-gate structure. We also observe that the electron density is mainly confined in the center of the silicon channel due to the ultrasmall channel thickness.

The LDOS and conduction band edge along the center of the silicon layer are illustrated in Fig. 15. The wave phenomenon is evident. In addition, compared with Fig. 6, the conduction band edge in the channel part is relatively flat. This suggests that the
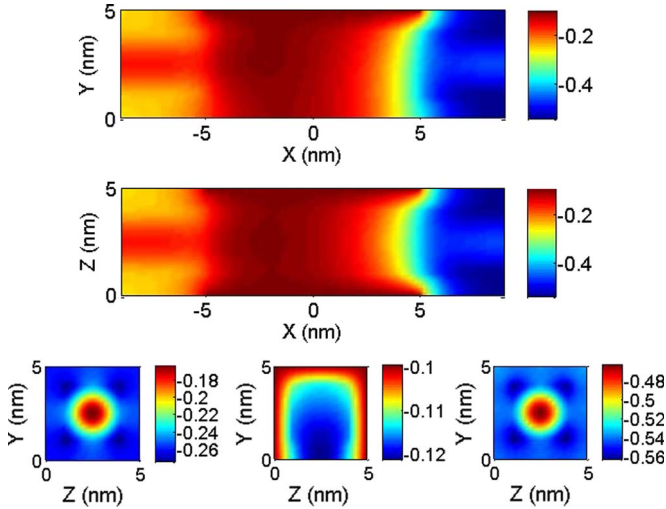
Fig. 13. Two-dimensional plot of the potential distribution in (top) XY plane, (middle) XZ plane, and (bottom left) YZ plane at the source end, (bottom middle) channel center, and (bottom right) drain end. $V_G = V_D = 0.3$ V.

Fig. 14. Two-dimensional plot of the electron distribution in (top) XY plane, (middle) XZ plane, and (bottom left) YZ plane at the source end, (bottom middle) channel center, and (bottom right) drain end. $V_G = V_D = 0.3$ V.
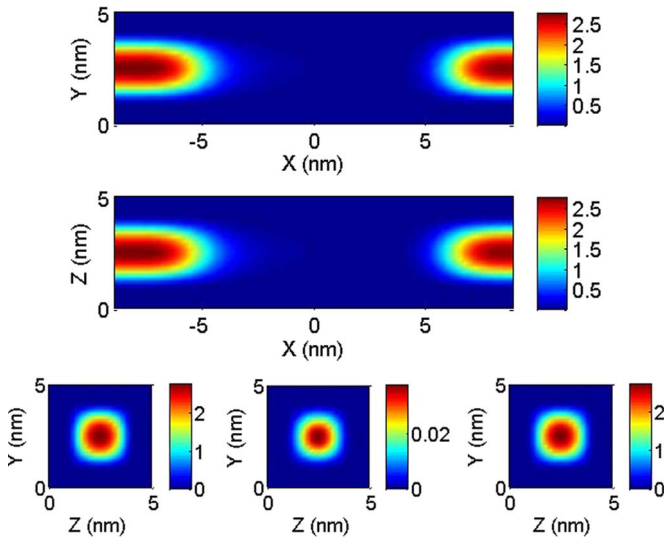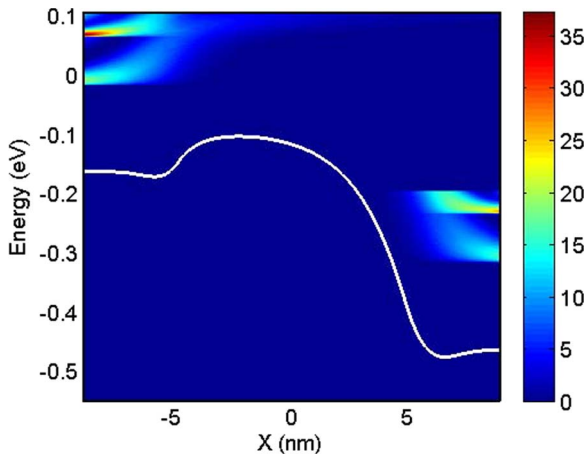

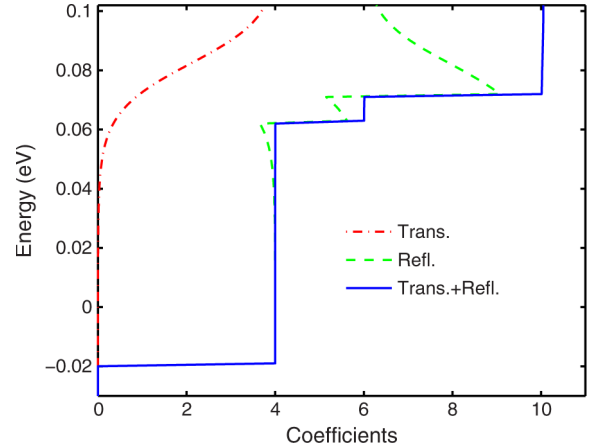Fig. 15. LDOS along the center of the silicon layer. $V_G = V_D = 0.3$ V.

Fig. 16. Transmission and reflection coefficients defined in the Landauer formula for the electrons coming from the source. $V_G = V_D = 0.3$ V.

TABLE II
LIST OF THE ENERGY POINTS, CPU TIME, AND CURRENT

	Energy Points	Current (nA)	CPU time (seconds)	Speed Up
Direct method	566	253.24	9122	1.0×
$\varepsilon = 2 \times 10^{-2}$, Order [3/3]	61	253.24	1277	7.1×
$\varepsilon = 2 \times 10^{-2}$, Order [4/4]	56	253.23	1270	7.2×
$\varepsilon = 2 \times 10^{-2}$, Order [5/5]	48	253.24	1163	7.8×
$\varepsilon = 4 \times 10^{-2}$, Order [3/3]	58	253.22	1166	7.8×
$\varepsilon = 4 \times 10^{-2}$, Order [4/4]	52	253.25	1141	8.0×
$\varepsilon = 4 \times 10^{-2}$, Order [5/5]	44	253.24	1076	8.5×

potential in the channel is mainly modulated by the gates, and the short-channel effect due to drain-induced barrier lowering is effectively suppressed.

Transmission and reflection coefficients are plotted in Fig. 16. Again, our approach produces very accurate results (note that the summation of transmission and reflection is an integer) over the entire energy band interested.

The comparison of various orders of Padé approximant with different CFH tolerances for this 3-D case is summarized in Table II (matrix solver is the same as the 2-D case). The reference is the direct method with energy grid 0.001 eV. The bias is $V_G = V_D = 0.3$ V. It is observed that the number of inversion points is reduced by over one order of magnitude with our method. The accuracy is mainly determined by the CFH tolerance. More accurate results can be obtained by minimizing the CFH tolerance but at the cost of more computer time since more expansion points are needed. To achieve the same accuracy, higher order Padé approximant takes less computational cost as it requires less expansion points. When $\varepsilon = 4 \times 10^{-2}$, the drain currents obtained are still very accurate, and it can be over eight times faster.

IV. CONCLUSION

In this paper, quantum ballistic transport of multiterminal devices has been modeled by the self-consistent Schrödinger–Poisson system. The AWE integrated with the CFH technique is proposed to accelerate the solution of the

Schrödinger equation in a wide energy range. Numerical results show that this method can reduce by over eight times the computer time, and the precision can be controlled to an acceptable level. The characteristic parameters of the device, such as the spectral density, LDOS, and transmission (reflection) coefficients at any energy are readily accessed by this method.

As a general method for wideband simulation, this algorithm can be incorporated to the finite-element method [2], [6] in a similar way. In addition, it can be combined with the coupled (uncoupled) mode space approach [3], [5] to further improve the efficiency. To extend this work to include inelastic scattering, however, is nontrivial, since a full NEGF simulation is required. The implementation of AWE to the Green's function is obviously more expensive, and the scattering self energy is not analytical in most cases. Extension to the tight-binding and the first principle model can be realized if the derivatives of the contact self energy can be obtained.

APPENDIX
DERIVATION OF THE MATRICES

Consider a typical case where there are two contacts (contact 1 and contact 2) and the directions ξ_α and η_α ($\alpha = 1, 2$) are the same as x and y. We apply a second-order central difference method to discretize the 2-D Schrödinger equation (1) by using the following formulas:

$$
\begin{aligned}
\nabla &\cdot \left[\frac{1}{m^*(x,y)} \nabla \psi \right]_{x=x_i, y=y_j} \\
&\approx \frac{1}{\Delta x^2} \left(\frac{\psi_{i+1,j} - \psi_{i,j}}{m^*_{i+1/2,j}} - \frac{\psi_{i,j} - \psi_{i-1,j}}{m^*_{i-1/2,j}} \right) \\
&\quad + \frac{1}{\Delta y^2} \left(\frac{\psi_{i,j+1} - \psi_{i,j}}{m^*_{i,j+1/2}} - \frac{\psi_{i,j} - \psi_{i,j-1}}{m^*_{i,j-1/2}} \right)
\end{aligned} \quad \text{(A.1)}
$$

where $\psi_{i,j} = \psi(x_i, y_j)$ and $m^*_{i\pm1/2, j\pm1/2} = m^*((x_i + x_{i\pm1})/2, (y_i + y_{i\pm1})/2)$, $i = 1, 2, \ldots, N_x, j = 1, 2, \ldots, N_y$. Δx and Δy are the uniform grid spacing in the x and y directions.

It is apparent that when $i = 1$ ($i = N_x$ is similar), we have to specify the values $\psi_{0,j}$ for $j = 1, 2, \ldots, N_y$. These values are directly obtained from the solution in contact 1 (15) as

$$
\begin{aligned}
\psi_{0,j} &= -2ia_n^1 \chi_n^1(y_j) \sin\left(k_n^1 \Delta x\right) \\
&\quad + \sum_{m=1}^{N_1} \left(\int \chi_m^1(y) \psi(x_1, y) dy \right) \chi_m^1(y_j) \exp(ik_m^1 \Delta x)
\end{aligned} \quad \text{(A.2)}
$$

$$
\begin{aligned}
&= -2ia_n^1 \chi_n^1(y_j) \sin\left(k_n^1 \Delta x\right) \\
&\quad + \sum_{m=1}^{N_1} \left(\sum_{j'=1}^{N_y} \chi_m^1(y_{j'}) \psi_{1,j'} \Delta y \right) \chi_m^1(y_j) \exp(ik_m^1 \Delta x)
\end{aligned} \quad \text{(A.3)}
$$

where the integration is replaced by a summation using trapezoid rule. Similarly, using the solution in contact 2, we can obtain the values $\psi_{N_x+1,j}$ for $j = 1, 2, \ldots, N_y$.

Writing the discretized equations with matrix form, we have (16). Matrix $\mathbf{H}$ for the isolated device is

$$
\mathbf{H} = \begin{bmatrix}
\mathbf{H}_1 & \mathbf{T}_{12}^\dagger & \mathbf{0} & \cdots & & \mathbf{0} \\
\mathbf{T}_{12} & \mathbf{H}_2 & \mathbf{T}_{23}^\dagger & \ddots & & \vdots \\
\mathbf{0} & \ddots & \ddots & \ddots & & \mathbf{0} \\
\vdots & \ddots & \mathbf{T}_{N-2,N-1} & \mathbf{H}_{N-1} & \mathbf{T}_{N-1,N}^\dagger \\
\mathbf{0} & \cdots & & \mathbf{0} & \mathbf{T}_{N-1,N} & \mathbf{H}_N
\end{bmatrix} \quad \text{(A.4)}
$$

where $\mathbf{H}_i$ is the tri-diagonal Hamiltonian matrix for layer i ($i = 1, 2, \ldots, N_x$), and $\mathbf{T}_{ij}$ is the diagonal matrix that represents the coupling between adjacent layers. Matrix $\mathbf{S}$ for self energy is

$$
\mathbf{S} = \begin{bmatrix}
\mathbf{S}^1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \vdots \\
\mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\
\vdots & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{S}^2
\end{bmatrix} \quad \text{(A.5)}
$$

where the nonzero elements are

$$
\mathbf{S}_{j,j'}^\alpha = -\frac{\hbar^2 \Delta y}{2m^*_{1/2,j} \Delta x^2} \sum_{m=1}^{N_\alpha} \chi_m^\alpha(y_{j'}) \chi_m^\alpha(y_j) \exp\left(ik_m^\alpha \Delta x\right),
$$

$$
\alpha \in \{1, 2\}, \text{ and } (j, j') \in \{1, \ldots, N_y\}. \quad \text{(A.6)}
$$

Vector $\mathbf{v}$ is

$$
\mathbf{v} = \begin{cases}
\begin{bmatrix} \mathbf{v}^1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T, & \text{wave from the left} \\
\begin{bmatrix} 0 & \cdots & 0 & 0 & \mathbf{v}^2 \end{bmatrix}^T, & \text{wave from the right}
\end{cases} \quad \text{(A.7)}
$$

where the nonzero elements are

$$
\mathbf{v}_j^\alpha = \frac{\hbar^2}{2m^*_{1/2,j} \Delta x^2} 2ia_n^\alpha \chi_n^\alpha(y_j) \sin\left(k_n^\alpha \Delta x\right),
$$

$$
\alpha \in \{1, 2\}, \text{ and } j \in \{1, \ldots, N_y\}. \quad \text{(A.8)}
$$

REFERENCES

[1] D. Vasileska and S. M. Goodnick, *Computational Electronics*. San Rafael, CA: Morgan & Claypool, 2006.
[2] C. S. Lent and D. J. Kirkner, "The quantum transmitting boundary method," *J. Appl. Phys.*, vol. 67, no. 10, pp. 6353–6359, May 1990.
[3] E. Polizzi and N. Ben Abdallah, "Subband decomposition approach for the simulation of quantum electron transport in nanostructures," *J. Comput. Phys.*, vol. 202, no. 1, pp. 150–180, Jan. 2005.
[4] C. D. Cheng and Q. H. Liu, "3D quantum transport solver based on the perfectly matched layer and spectral element methods for the simulation of semiconductor nanodevices," *J. Comput. Phys.*, vol. 227, no. 1, pp. 455–471, Nov. 2007.
[5] J. Wang, E. Polizzi, and M. Lundstrom, "A three-dimensional quantum simulation of silicon nanowire transistors with the effective-mass approximation," *J. Appl. Phys.*, vol. 96, no. 4, pp. 2192–2203, Aug. 2004.
[6] P. Havu and V. Havu, "Nonequilibrium electron transport in two-dimensional nanostructures modeled using Green's functions and the

finite-element method," *Phys. Rev. B, Condens. Matter*, vol. 69, no. 11, p. 115 325, Mar. 2004.

[7] H. Y. Jiang, S. H. Shao, and W. Cai, "Boundary treatments in non-equilibrium Green's function (NEGF) methods for quantum transport in nano-MOSFETs," *J. Comput. Phys.*, vol. 227, no. 13, pp. 6553–6573, Jun. 2008.

[8] J. A. Torres and J. J. Saenz, "Improved generalized scattering matrix method: Conduction through ballistic nanowires," *J. Phys. Soc. Jpn.*, vol. 73, no. 8, pp. 2182–2193, Aug. 2004.

[9] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *J. Appl. Phys.*, vol. 91, no. 4, pp. 2343–2354, Feb. 2002.

[10] D. Mamaluy, D. Vasileska, M. Sabathil, T. Zibold, and P. Vogl, "Contact block reduction method for ballistic transport and carrier densities of open nanostructures," *Phys. Rev. B, Condens. Matter*, vol. 71, no. 24, p. 245 321, Jun. 2005.

[11] H. R. Khan, D. Mamaluy, and D. Vasileska, "Quantum transport simulation of experimentally fabricated nano-FinFET," *IEEE Trans. Electron Devices*, vol. 54, no. 4, pp. 784–796, Apr. 2007.

[12] G. Mil'nikov, N. Mori, Y. Kamakura, and T. Ezaki, "R-matrix theory of quantum transport and recursive propagation method for device simulations," *J. Appl. Phys.*, vol. 104, no. 4, p. 044 506, Aug. 2008.

[13] G. Mil'nikov, N. Mori, and Y. Kamakura, "R-matrix method for quantum transport simulations in discrete systems," *Phys. Rev. B, Condens. Matter*, vol. 79, no. 23, p. 235 337, Jun. 2009.

[14] G. Mil'nikov, N. Mori, and Y. Kamakura, "Application of the R-matrix method in quantum transport simulations," *J. Comput. Electron.*, vol. 9, no. 3/4, pp. 256–261, Dec. 2010.

[15] W. C. Chew, J. M. Jin, E. Michielssen, and J. Song, *Fast and Efficient Algorithms in Computational Electromagnetics*. Norwood, MA: Artech House, 2001.

[16] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[17] Z. Ren, "Nanoscale MOSFET: Physics, simulation and design," Ph.D. dissertation, Purdue Univ., West Lafayette, IN, 2001.

[18] Z. Ren, S. Goasguen, A. Matsudaira, S. S. Ahmed, K. Cantley, Y. Liu, M. Lundstrom, and X. Wang, *NanoMOS*, 2009, DOI: 10254/nanohub-r1305.11.

[19] N. Neophytou, A. Paul, M. S. Lundstrom, and G. Klimeck, "Simulations of nanowire transistors: Atomistic vs. effective mass models," *J. Comput. Electron.*, vol. 7, no. 3, pp. 363–366, Sep. 2008.
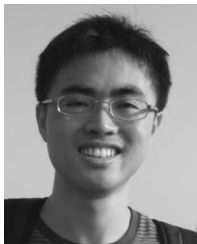
Jun Z. Huang received the B.S. degree in Electronic Science & Technology from Nan Kai University, Tianjin, China, in 2004, and the M.S. degree in electromagnetic and microwave technology from Shanghai Jiao Tong University, Shanghai, China, in 2010. He is currently working toward the Ph.D. degree with The University of Hong Kong, Pokfulam, Hong Kong.

From December 2008 to November 2009, he was a Full-Time Research Assistant with The University of Hong Kong, Pokfulam, Hong Kong. His research interests include modeling nanoscale transistors and computational electronics.

Weng Cho Chew (S'79–M'80–SM'86–F'93) received all his academic degrees from the Massachusetts Institute of Technology, Cambridge.

He worked on fast algorithms in computational electromagnetics in the last 20 years. From 1981 to 1985, he worked with Schlumberger-Doll Research as a Research Staff, a Program Leader, and a Manager. In 1985, He joined the University of Illinois at Urbana–Champaign (UIUC), Urbana–Champaign, and from 1995 to 2007, he was a Director with the Electromagnetics Laboratory, UIUC. From 2000 to 2005, he was a Founder Professor with UIUC, and from 2005 to 2009, he was the First Y. T. Lo Endowed Chair Professor. In 2006, he was a Cheng Tsang Man Professor with Nanyang Technological University, Singapore. From 2007 to 2011, he was the Dean of Engineering with The University of Hong Kong, Pokulam, Hong Kong. He is the author and coauthor of three books, over 350 journal papers, and over 400 conference papers.

Dr. Chew is a Fellow of OSA and IOP. He is currently the Editor-in-Chief of JEMWA/PIER journals. He is an ISI highly cited author (top 0.5). He was the recipient of numerous best paper and teaching awards and the CT Tai Distinguished Educator Award from IEEE AP-S in 2008.

Min Tang (M'09) received the B.S. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 2001; the M.S. degree in electrical engineering from Xi'an Jiao Tong University, Xi'an, China, in 2004; and the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2007.

He is currently a Lecturer with the Department of Electronic Engineering, Shanghai Jiao Tong University. His research interests include modeling and simulation of high-speed interconnects, computer-aided design of very large scale integrated circuits, and computational electromagnetics.

Lijun Jiang (S'01–M'04) received the B.S. degree in electrical engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1993; the M.S. degree from Tsinghua University, Beijing, in 1996; and the Ph.D. degree from the University of Illinois at Urbana–Champaign, Urbana–Champaign, in 2004.

From 1996 to 1999, he was an Application Engineer with the Hewlett-Packard Company. Since 2004, he has been a Postdoctoral Researcher, a Research Staff Member, and a Senior Engineer with IBM T.J. Watson Research Center, Yorktown Heights, NY. Since the end of 2009, he has also been an Associate Professor with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokulam, Hong Kong. His research interests include electromagnetics, EMC/EMI, antennas, multidisciplinary EDA solutions, RF and microwave technologies, high-performance computing, etc.

Dr. Jiang was the Semiconductor Research Cooperation (SRC) Industrial Liaison for several academic projects. Since 2009, he has been the SRC Packaging High-Frequency Topic TT Chair. He has served as the Technical Committee Member for IEEE EDAPS since 2010, the Scientific Committee Member of 2010 IEEE SMEE, and the Associate Guest Editor of IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION since 2011. He also serves as the reviewer of many primary electromagnetics journals and special sessions organizers for many international conferences. He was a recipient of the HP Star Award in 1998, the IEEE MTT Graduate Fellowship Award in 2003, the Y. T. Lo Outstanding Research Award in 2004, and the IBM Research Technical Achievement Award in 2008.