

Electron-only explicit screening quantum transport model for semiconductor nanodevices

Yuanchen Chu^{*†‡}, Prasad Sarangapani^{†‡}, James Charles^{†‡}, Gerhard Klimeck^{‡§} and Tillmann Kubis^{‡§}

*Email: chu72@purdue.edu

[†]School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, 47907, USA

[‡]Network for Computational Nanotechnology, Purdue University, West Lafayette, Indiana, 47907, USA

[§]Center for Predictive Materials and Devices, Purdue University, West Lafayette, Indiana, 47907, USA

Abstract—State of the art quantum transport models for semiconductor nanodevices attribute negative (positive) unit charges to states of the conduction (valence) band. Hybrid states that enable band-to-band tunneling are subject to interpolation that yield model dependent charge contributions. In any nanodevice structure, these models rely on device and physics specific input for the dielectric constants. This work exemplifies the large variability of different charge interpretation models when applied to ultrathin body transistor performance predictions. To solve this modeling challenge, an electron-only band structure model is extended to atomistic quantum transport. Performance predictions of MOSFETs and tunneling FETs confirm the generality of the new model and its independence of additional screening models.

Index Terms—Electron-only, tunneling FET, ultrathin body transistor, NEGF

I. INTRODUCTION

As the scaling of Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) has reached sub-10 nm regime, power consumption has become a major concern. The advantages of lowering the dynamic power consumption by reducing the supply voltage are fast disappearing as the static power has begun to dominate due to the exponential increase of the subthreshold leakage current [1]. Band-to-band tunneling field-effect transistor (TFET) is among the most promising candidates for future integrated circuits (ICs) due to its ability to beat the 60mV/decade limit of the subthreshold swing (SS) [2]. Having a smaller subthreshold swing enables a reduction of both the supply voltage and the subthreshold leakage current, thus further lowering the power consumption of the ICs.

The quantitative prediction of TFET performance requires self-consistent solutions of charge distributions and quantum transport equations. The nonequilibrium Greens Function (NEGF) method is widely accepted for this purpose [3]. The standard model (termed as excess-charge approach or ECA) to interpret the particle density in quantum transport calculations distinguishes the charge carrier type: An electron (hole) in conduction (valence) band state of n-type (p-type) MOSFET is considered to contribute a negative (positive) unit charge. This concept limits the computational load to solving electrons (holes) in the conduction (valence) band only, i.e. a few $k_B T$ of energy in addition to the energy range spanned by the

applied bias voltage. In the Poisson equation, the dielectric constant is then typically set to the materials constant. In tunneling devices, a particle with energy above (below) the conduction (valence) band is still considered to contribute a negative (positive) unit charge. For energies between the conduction and the valence band, various charge interpolation schemes exist.

As will be shown in detail in this paper, the standard model fails in various ways, both for conventional as well as for band-to-band tunneling transistors. It turns out - consistent with previous findings [4] in literature - that the electrostatic screening of valence band electrons that do not take part in transport is device physics (and not only material) dependent. The charge interpolation schemes required for band-to-band tunneling devices host an arbitrariness that severely limits the reliability of device performance predictions. Any such interpolation also suffers from incompatibility with the NEGF method as discussed in detail in [5]. This is consistent with previous findings in broken-gap optoelectronic bandstructure calculations [6].

To solve these modeling challenges, a numerically efficient charge self-consistent model is developed in the NEGF formalism where electrons are considered throughout all bands including the deepest lying valence band. Performance predictions of MOSFETs and TFETs confirm the generality of the new model and its independence of additional screening models.

II. METHODS

In band-to-band tunneling (BTBT) situations, both valence band and conduction band states have to be considered, since states exist that overlap with conduction and valence band simultaneously. The density of such states is translated with a heuristic interpolation factor λ into their charge density contribution. The expression for the charge density contribution of each individual lead is given by

$$\begin{aligned} -qn_{ECA}(E, k) &= qp(E, k) - qn(E, k) \\ &= q [1 - f(E, \mu)] \lambda(E, k) |\psi(E, k)|^2 \\ &\quad - qf(E, \mu) [1 - \lambda(E, k)] |\psi(E, k)|^2, \end{aligned} \quad (1)$$

where E is the state energy, $\psi(E, k)$ is the injected conduction or valence band state, μ is the lead Fermi level and q is the

TABLE I
THREE HEURISTIC INTERPOLATION MODELS

Label	α	$E \leq D$	$E > D$
A	0.2	$\lambda = 1 - F(E_V)$	$\lambda = F(E_C)$
B	0.5	$\lambda = 1 - F(E_V)$	$\lambda = F(E_C)$
C	0.5	1	0

positive unit charge. The factor λ heuristically interpolates between the positive and negative charge interpretations of valence and conduction band. Therefore, λ depends on the electrostatic potential, the conduction and valence band edges and the chosen heuristic interpolation model. In general λ is a function of energy, in-plane momentum and position. The interpolation factor $\lambda(E)$ is applied when the energy is above the valence band edge and below the conduction band edge. Below (above) the valence (conduction) band edge, $\lambda(E)$ is equal 1 (0). If transport happens exclusively in valence or conduction bands (e.g. in MOSFETs), no interpolation factor is applied in ECA. In this paper, three commonly used [7], [8] heuristic models for λ have been chosen as representatives (summarized in Table. I). All three heuristic models distinguish the interpolation factor λ for energies E below and above a delimiter energy D given by

$$D = (1 - \alpha)E_V + \alpha E_C, \quad (2)$$

where α is a unitless number and E_V (E_C) is the valence (conduction) band edge. For simplicity, a function $F(\mathcal{E})$ is defined as

$$F(\mathcal{E}) = \frac{\mathcal{E} - E}{2(\mathcal{E} - D)}, \quad (3)$$

For energies above D , λ equals the function F evaluated for $\mathcal{E} = E_C$ and $\lambda = 1 - F(E_V)$ otherwise. The space charge density of the Poisson equation is obtained by summing the electron or hole charge density with the background doping density.

As a consistent alternative to the ECA model, this work extends the charge self-consistent model (termed as full-band approach or FBA) of Refs [6] and [9] to atomistic quantum transport of band-to-band tunneling devices within the NEGF formalism. Every state solved within the quantum transport method is considered electronic and contributes, if occupied, a negative unit charge. This is irrespective of which band that state is in. However, this model requires resolving the density contribution of all occupied states.

The standard NEGF treatment of density calculations requires integrating the diagonal of the retarded Greensfunction GR over an energy interval that covers all occupied states. The recursive Greens function implementation of NEMO5 is applied to solve for the diagonal of GR. Greens functions and self-energies are matrices in the position space indicated in bold font. Most of the following equations involve the diagonal of the Greens functions only which is denoted in nonbold letters.

The total electron density n_{FBA} is separated into an equilibrium n_{eq} and a nonequilibrium part n_{neq} .

$$n_{FBA} = n_{eq} + n_{neq}. \quad (4)$$

The equilibrium electron density contribution is dependent on one contact Fermi function (e.g. the left one) and is given by

$$n_{eq} = \sum_k \int_{-\infty}^{\infty} \frac{-\text{Im}[G^R(E, k)]}{\pi} f_L(E, \mu_L) dE, \quad (5)$$

where μ_L is the Fermi level of the left contact. Many atomistic models yield 10s of eV with hundreds of van Hove singularities of fully occupied valence bands [10], which are all considered within n_{eq} . To avoid resolving all these states on a real energy mesh which poses immense numerical loads, n_{eq} is solved with the Residual theorem

$$n_{eq} = \sum_k \left\{ \int_{H+C} \frac{\text{Im}[G^R(E, k)]}{\pi} f_L(E, \mu_L) dE + i2k_B T \sum_{pole} G^R(E_{pole}, k) \right\}. \quad (6)$$

The poles of the integrand originate from the Fermi function of the (left) contact. These poles are located at $E_{pole} = \mu_L + ik_B T \pi(2m + 1)$ with $\text{Res}(E_{pole}) = -k_B T$, $m \in \mathbb{N}$. A typical integration contour [9] is shown in Fig. 1. The integration contour consists of a semicircular part (C) whose lower bound is set about 1 eV below the lowest eigenvalue of the system. The horizontal part of the contour (H) is parallel to the real energy axis. The maximum real part of H is exceeding μ_L by 25 $k_B T$ to include the complete tail of the contact Fermi function in the density calculation. The small contour portion that closes the integration contour beyond the horizontal section H does not have a net contribution to Eq. (5). When the imaginary part of the integration contour is large enough (such as indicated in Fig. 1), numerical solutions of Eq. (6) converge with few tens of contour points.

The integral of the non-equilibrium electron density must be performed along the real energy axis since the integrand is not analytic in the entire complex plane.

$$n_{neq} = \sum_k \int_{-\infty}^{\infty} \text{diag} \left\{ \mathbf{G}^R(\mathbf{E}, \mathbf{k}) \frac{\text{Im}[\boldsymbol{\Sigma}^R(\mathbf{E}, \mathbf{k})]}{\pi} \mathbf{G}^R(\mathbf{E}, \mathbf{k})^\dagger \right\} \times [f_L(E, \mu_L) - f_R(E, \mu_R)] dE. \quad (7)$$

Equation (7) involves a matrix product of Green's functions and a self-energy indicated by bold letters. The integration window is restricted by the two contact Fermi functions (μ_R being the right contact's Fermi level) and is approximately the same energy window considered in the ECA. Compared to the ECA, the extra computational load in FBA is given by a few tens of energy points for the integral of the equilibrium electron density in Eq. (6). This is a negligible addition given the hundreds or thousands of energy points typically needed to resolved the non-equilibrium density contribution n_{neq} .

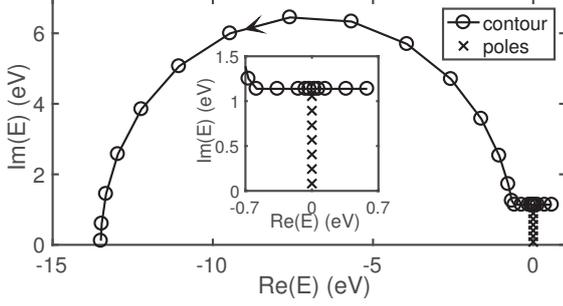


Fig. 1. A typical integration contour used for n_{eq} in Eq. (4). The Fermi level of the left contact μ_L is set to 0. Poles enclosed by the contour are marked by crosses and highlighted in the inset. The inset also illustrates the dense distribution of mesh points around the Fermi level that ensure a well resolved contact Fermi function. The arrow on the contour indicates the direction of the integral in the complex plane.

In charge self-consistent FBA calculations, the Poisson equation requires a positive background charge n_{core} to allow the presence of electrons in the devices. This background charge is assumed to completely compensate the electronic charge density of the respective undoped device in equilibrium. The total space charge ρ_{FBA} is given by the sum of n_{eq} , n_{neq} , n_{core} and the doping density n_{doping} .

$$\rho_{FBA} = -qn_{FBA} + qn_{core} + qn_{doping}, \quad (8)$$

The definition of n_{core} and Eq. (8) approximate Madelung-like potentials due to charge transfer of e.g. center-device atoms and surface atoms into one reference situation. Therefore, the density ρ_{FBA} accounts for deviations of the charge density vs. the reference situation of the undoped and equilibrated device. The dielectric constant of vacuum is used in the Poisson equation for the semiconductor materials in FBA since the screening of all valence band electrons is explicitly included in the calculation. This makes the FBA model independent of the material and device specific dielectric screening.

RESULTS

A. Convergence behavior of n_{eq} calculation

The numerical convergence of solving the equilibrium electron density n_{eq} with Eqs. (5) and (6) with varying number of energy points per momentum point is compared for homogeneous 3D silicon in Figs. 2 (a) and (b). In both cases, the electronic Brillouin zone is resolved with 225 momentum points. The equilibrium density of Eq. (6) converges with only a few complex energy points (contour points and poles) per momentum point. In contrast, many and hard to resolve van Hove singularities on the real-energy axis prevent Eq. (5) to fully converge even with an immense number of energy points.

B. Transfer characteristics of silicon MOSFETs

Transfer characteristics resulting from FBA and ECA charge self-consistent calculations of a p-type and a n-type silicon

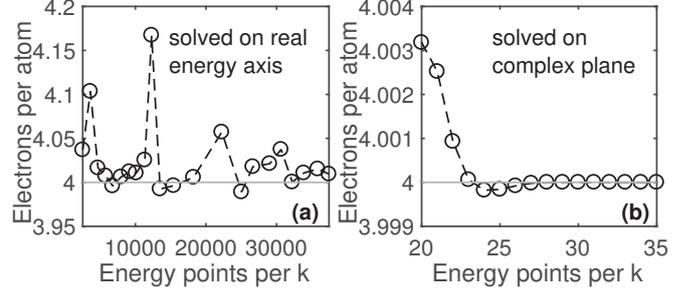


Fig. 2. The atom resolved electron density n_{eq} vs number of energy points per momentum point for the real energy integral in Eq. (5) (a) and for the complex contour integral of Eq. (6) (b). The converged number of 4 valence Si electrons is indicated with a solid line to guide the eye. The n_{eq} converges with only 30 energy points in Eq. (6) while no convergence is observed when solved along the real energy axis even with an immense number of energy points.

ultra-thin body double-gate MOSFET are shown in Fig. 4. Both MOSFET devices have a structure as shown in Fig. 3, with body thickness, channel and source/drain lengths of 1.6 nm, 10.8 nm and 11.4 nm, respectively. A doping concentration of $1 \times 10^{20} \text{ cm}^{-3}$ is assumed in the source and drain regions of both transistor types. A drain-to-source voltage V_{DS} of 0.4 V is applied. The equivalent oxide thickness (EOT) of top and bottom oxides is 1 nm. It is reported both experimentally [4] and theoretically [11] that the dielectric constant of silicon ultra-thin films reduces with the film thickness. A dielectric constant of 9.9 is used in the ECA simulations following Ref. [11]. In FBA calculations, only the vacuum dielectric constant enters the Poisson equation.

Results of both models, for the transfer characteristics and band profiles agree very well. This is particularly true compared to TFET situation (see the following subsection) that defines the common ECA/FBA difference scale. The good

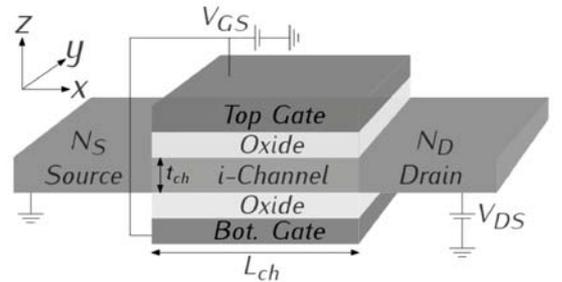


Fig. 3. Schematic of the ultra-thin body double-gate transistors used in all transfer characteristic predictions of this work. The gate bias is controlled by V_{GS} . Electron transport occurs along x direction when a non-zero V_{DS} is applied. The channel of the device is confined along z direction and t_{ch} is the channel thickness. Periodic boundary condition is assumed along y direction. L_{ch} is the channel (gate) length. N_S and N_D are doping concentrations in source and drain regions, respectively.

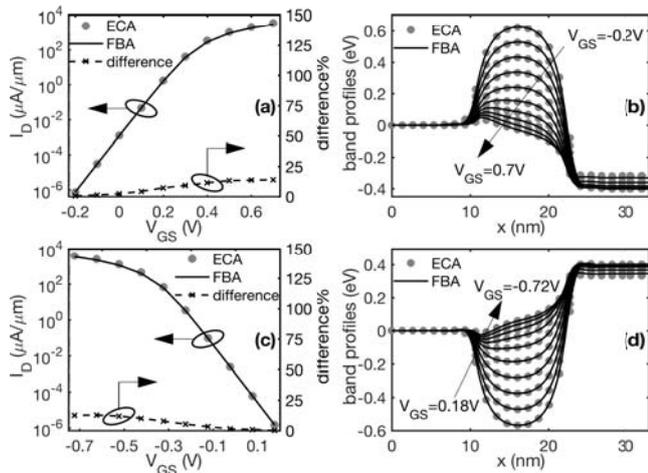


Fig. 4. Transfer characteristics $I_D - V_{GS}$ of silicon ultra-thin body double-gate nMOSFET (a) and pMOSFET (c) of Fig. 3 predicted with ECA and FBA. The percentage difference is plotted in dashed lines with cross markers. (b) and (d) Band profiles of ECA and FBA corresponding to V_{GS} nodes in (a) and (c), respectively.

agreement is expected since electronlike and holelike states are clearly separated in these devices. The maximum relative difference of the drain current in the two models is below 15% and the maximum potential difference is below $k_B T$. Note that the dielectric constant of the ECA model could serve as a fitting degree of freedom to match the FBA results. This finding emphasizes the strength of the FBA model to explicitly handle the electrostatic response of the deep lying valence electrons with a marginal increase in computational cost.

C. Silicon TFET

To compare the predictions of ECA and FBA for BTBT devices, the device structure of Fig. 3 is considered with the source (drain) doping being p-type (n-type). The source and drain doping concentrations are $5 \times 10^{19} \text{ cm}^{-3}$ and $2 \times 10^{19} \text{ cm}^{-3}$, respectively. The body thickness, channel and source/drain lengths are 1.6 nm, 10.8 nm and 11.4 nm, respectively. A dielectric constant of 9.9 is used in the ECA model. The ECA utilizes heuristic models to distinguish electronlike and holelike charge contributions in BTBT devices. Three commonly used heuristic models (summarized in Table. I) are applied in ECA and results are presented and compared to the FBA result (see Fig. 5).

The difference in the performance predictions of ECA and FBA can be understood from Fig. 6. The electron density in the bandgap (at around 10 nm) is considered as electrons in FBA, whereas in ECA, a charge prefactor is assigned to it. This prefactor depends on the position of the hole/electron delimiter and the considered interpolation scheme (see Table. I). A snapshot of the delimiter for $\alpha = 0.5$ is illustrated in Fig. 6. Consequently, that prefactor differs in the three applied heuristic models. The different prefactors in turn impact the interpretation of ECA charge and the electrostatic potential around the tunnel junction. Thus, the TFET transfer

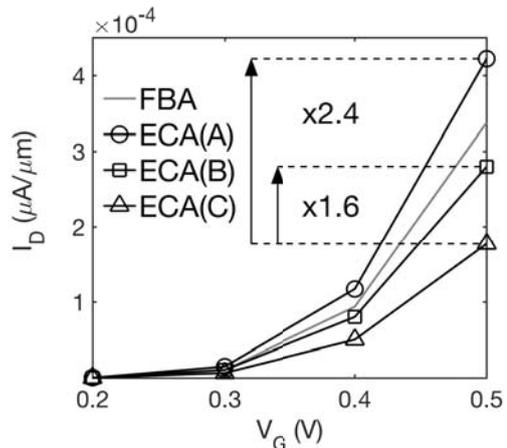


Fig. 5. Transfer characteristics $I_D - V_{GS}$ at $V_{DS} = 0.3 \text{ V}$ of a silicon ultra-thin body double-gate TFET. Results of FBA and ECA with three different heuristic models are shown. The maximum deviation of the three ECA results relative to their average is plotted in dashed line with cross markers.

characteristic prediction is sensitive to the chosen ECA model. This is indicated by the large deviations in Fig. 5 among the three considered ECA models.

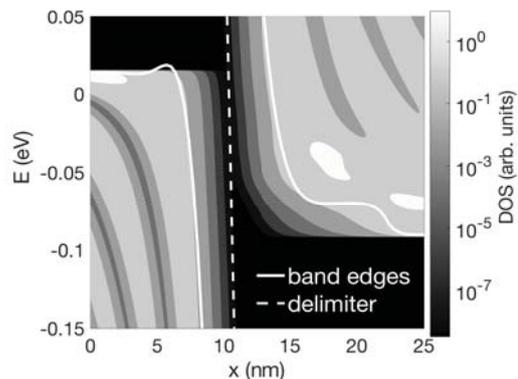


Fig. 6. Contour plot of the energy and position resolved density of states of the silicon TFET simulated in Fig. 5 at $k = \Gamma$. The conduction and valence band edges are represented in white solid lines. The white dashed line depicts a hole/electron delimiter for $\alpha = 0.5$ in the band gap which is used to distinguish electron and hole states in ECA.

CONCLUSION

We have developed a charge self-consistent model for quantum transport calculations in TFETs where standard charge self-consistent approaches that distinguish between electrons and holes fail. Since the new model considers all valence electrons explicitly, no extra input of screening constants is required. The new model provides a much wider application space than the conventional electron-hole picture with a marginal increase in computational cost.

REFERENCES

- [1] J. Kao, S. Narendra, and A. Chandrakasan, "Subthreshold leakage modeling and reduction techniques," in *Proceedings of the 2002 IEEE/ACM International Conference on Computer-aided Design*, ser.

- ICCAD '02. New York, NY, USA: ACM, 2002, pp. 141–148. [Online]. Available: <http://doi.acm.org/10.1145/774572.774593>
- [2] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, “Band-to-band tunneling in carbon nanotube field-effect transistors,” *Phys. Rev. Lett.*, vol. 93, p. 196805, Nov 2004. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.93.196805>
- [3] A. Mátyás, T. Kubis, P. Lugli, and C. Jirauschek, “Comparison between semiclassical and full quantum transport analysis of THz quantum cascade lasers,” *Physica E: Low-Dimensional Systems and Nanostructures*, vol. 42, no. 10, pp. 2628–2631, 2010.
- [4] H. G. Yoo and P. M. Fauchet, “Dielectric constant reduction in silicon nanostructures,” *Phys. Rev. B*, vol. 77, p. 115355, Mar 2008. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.77.115355>
- [5] Y. Chu, P. Sarangapani, J. Charles, G. Klimeck, and T. Kubis, “Explicit screening full band quantum transport model for semiconductor nanodevices,” *Journal of Applied Physics*, vol. 123, no. 24, p. 244501, 2018. [Online]. Available: <https://doi.org/10.1063/1.5031461>
- [6] T. Andlauer and P. Vogl, “Full-band envelope-function approach for type-II broken-gap superlattices,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 80, no. 3, pp. 1–7, 2009.
- [7] M. Luisier and G. Klimeck, “Atomistic full-band design study of inas band-to-band tunneling field-effect transistors,” *IEEE Electron Device Letters*, vol. 30, no. 6, pp. 602–604, June 2009.
- [8] P. Long, E. Wilson, J. Z. Huang, G. Klimeck, M. J. W. Rodwell, and M. Povolotskyi, “Design and simulation of gasb/inas 2d transmission-enhanced tunneling fets,” *IEEE Electron Device Letters*, vol. 37, no. 1, pp. 107–110, Jan 2016.
- [9] D. A. Areshkin and B. K. Nikolić, “Electron density and transport in top-gated graphene nanoribbon devices: First-principles Green function algorithms for systems containing a large number of atoms,” *Physical Review B - Condensed Matter and Materials Physics*, vol. 81, no. 15, pp. 1–17, 2010.
- [10] Y. P. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, “Tight-binding analysis of si and gaas ultrathin bodies with subatomic wave-function resolution,” *Phys. Rev. B*, vol. 92, p. 085301, Aug 2015. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevB.92.085301>
- [11] S. Markov, G. Penazzi, Y. Kwok, A. Pecchia, B. Aradi, T. Frauenheim, and G. Chen, “Permittivity of Oxidized Ultra-Thin Silicon Films from Atomistic Simulations,” *IEEE Electron Device Letters*, vol. 36, no. 10, pp. 1076–1078, 2015.