Essential Physics of Carrier Transport in Nanoscale MOSFETs

Mark Lundstrom, Zhibin Ren, and Supriyo Datta
Electrical and Computer Engineering, Purdue University
West Lafayette, IN 47907

Abstract – A simple, physical view of carrier transport in nanoscale MOSFETs is presented. The role of ballistic transport, scattering and off-equilibrium transport, and quantum transport are illustrated by numerical simulation, and the limitations of common approaches used for device TCAD are examined.

I. INTRODUCTION

Recent work shows that MOSFETs now operate surprisingly close to their ballistic limits [1, 2]. In the past, physically detailed simulations have led us to an understanding of the hot carrier and off-equilibrium transport effects such as velocity overshoot that occur under collision-dominated conditions [3-5], but to develop nanoscale MOSFETs and sound, physics-based models for them, a clear understanding of quasi-ballistic transport is needed. In this paper we show that the essential transport physics as illuminated by detailed simulations, can be understood in a simple way, one that helps to interpret detailed simulations (and even to detect errors in them) and that should prove useful in developing physically sound models for nanoscale MOSFET's.

Figure 1 summarizes the essential physical picture. Carriers are injected into the channel from a thermal equilibrium resevoir (the source), across a potential energy barrier whose height is modulated by the gate voltage, into the channel, which begins at the top of the barrier. The beginning of the channel is populated by carriers injected from the thermal equilibrium source. The density of carriers at the top of the barrier is set by MOS electrostatics so that the charge in the semiconductor balances that in the gate. For a device with low DIBL, equilibrium, 1D MOS electrostatics apply at this point, so the inversion layer density may be computed as a 1D MOS capacitor. Above threshold,

$$q n_s(0) \approx C_{ox}(V_{GS} - V_T), \tag{1}$$

where C_{ox} is the effective oxide capacitance (as influenced by quantum mechanical confinement, polysilicon depletion, etc.).

Some carriers injected into the channel backscatter and return to the source; others flow out the drain and comprise the steady-state drain current, I_D. Assuming current continuity, I_D may be evaluated at the beginning of the channel where the carrier density is known to find

$$I_D = W C_{ox}\langle \upsilon(0)\rangle (V_{GS} - V_T), \tag{2}$$

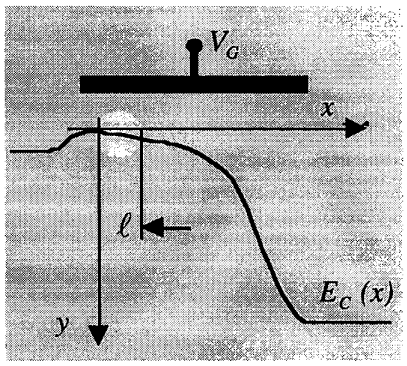

Fig. 1 The conduction band edge vs. position from the source to the drain of a nanoscale MOSFET under high gate and drain bias.

where $\langle \upsilon(0) \rangle$ is the average velocity of carriers at the beginning of the channel. The maximum value of $\langle \upsilon(0) \rangle$ is approximately the uni-directional thermal velocity, υ_T, because the positive velocity carriers at the beginning of the channel were injected from the thermal equilibrium source [6, 7]. Backscattering from the channel determines how close to this upper limit the device operates. Velocity overshoot occurs within the channel and determines the carrier density profile, which, through Poisson's equation sets the self-consistent profile through out the entire channel.

Figure 2 is a fluid flow analogy for the MOSFET under high gate and drain bias. Carrier transport through the drain end of the channel is rapid, because strong velocity overshoot occurs there. As a result, the d.c. current is controlled by how rapidly carriers are transported across a short low-field region near the beginning of the channel. Carriers diffuse across the beginning of the channel much the same way that they diffuse across the base of a bipolar transistor, and they are collected by the high-field portion of the channel much as in the collector of a bipolar transistor. The length of the current-limiting region at the beginning of the channel (which is set by 2D electrostatics as influenced by velocity overshoot within the channel) is about one-mean-free path, which means that a successful transport model must describe transport under quasi-ballistic conditions. Most of the models in use today, however, were developed under collision-dominated assumptions. These models often fail to describe quasi-ballistic transport in the current-limiting region and can, therefore, incorrectly predict MOSFET drain currents [8].

0-7803-6279-9/00/$10.00 © 2000 IEEE

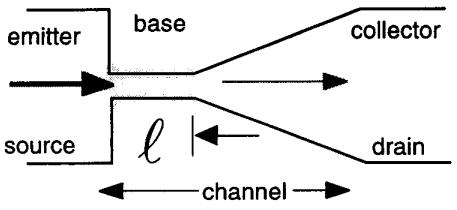

Fig. 2 A fluid flow analogy for the MOSFET under high gate and drain bias conditions.

In the following sections, we use detailed, numerical simulations to confirm this basic physical picture and to expand upon it.

II. BALLISTIC TRANSPORT

The transport physics of nanotransistors most is clearly examined in the ballistic limit, but since current day devices operate so close to the ballistic limit, there is also a practical motivation to examine this limit. Ballistic *I-V* curves may be computed in two different ways. The first uses 1D calculations from Schrödinger-Poisson simulations [7]. The second combines 2D electrostatics, which is essential for MOSFETs, and a 1D transport model, which can be justified for thin-body SOI or for the inversion layer of a MOSFET. Figure 3 shows the self-consistent conduction band profile computed for an L = 10nm double gate SOI MOSFET

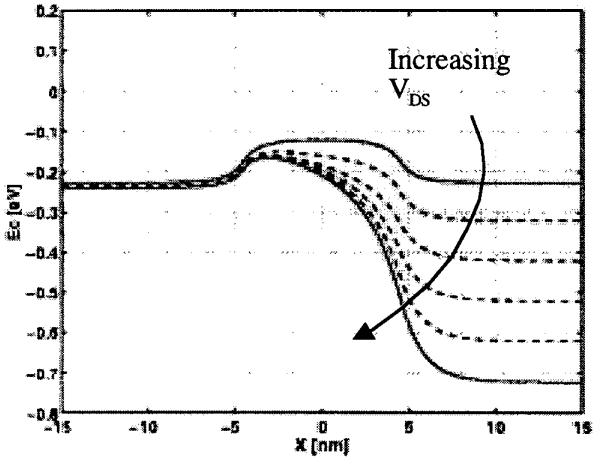

Fig. 3 The computed self-consistent conduction band edge vs. position for an L = 10 nm double gate MOSFET with t_{ox} = 1.0 nm, t_{Si} = 2.0 nm, and $V_G - V_T$ = 0.4V. Results are shown for several different drain biases.

In Sec. I, we argued that the average carrier velocity at the beginning of the channel was the uni-directional thermal velocity, v_T. Assuming that only one subband is occupied [7, 9],

$$v_T = \sqrt{\frac{2k_BT_L}{\pi m_t^*}}\left\{\frac{\mathcal{F}_{1/2}(\eta)}{\ell n(1+e^{\eta})}\right\}, \tag{3}$$

where $\eta = (E_F - E_C)/k_BT$, and the factor in brackets accounts for carrier degeneracy and approaches unity for a nondegenerate gas. (More generally, when multiple subbands are occupied, Schrödinger-Poisson simulations are needed.) Figure 4 shows the v_T vs. n_S characteristic for the 10 nm DG SOI MOSFET. Note that in subthreshold, $v_T \approx 1.2\times10^7$ cm/s, but that above threshold, the carriers become degenerate and the thermal injection velocity increases. These are the maximum velocities that can be observed at the source; they exceed the saturated velocity, but the origin of this high velocity is much different than that of conventional velocity overshoot that occurs in steep electric field gradients. (Nevertheless, there must be strong conventional velocity overshoot within the channel, or these high velocities will not be observed near the source.)

Figure 5 is a plot of $\langle v(0)\rangle$ vs. drain bias. Under low bias, the average velocity is nearly zero because the negative velocities of carriers injected from the drain nearly cancel the positive velocities of those injected from the source. When the drain bias exceeds a few kT/q, then the negative velocity carriers are suppressed, and the average velocity approaches v_T. Note that under a high drain bias, the carrier distribution at x = 0 has a highly nonequilibrium, hemi-Maxwellian, shape. This is also illustrated in Fig. 5, which shows the ratio, r, of the negative flux to the positive flux. Note that the net velocity saturates at $v_T = 1.7\times10^7$ cm/s when the drain bias is large enough to suppress the injection of negative-velocity carriers from the drain. This value is just the thermal injection velocity shown in Fig. 4.

For low V_{DS}, the velocity distribution is nearly symmetrical about v_x = 0, but under high drain bias, the distribution at the source assumes an asymmetrical, hemi-Maxwellian shape. It might appear, therefore, that $n_s(0)$ would be one-half of its equilibrium value, eq. (1). MOS electrostatics, however, demand that charge balance, so the conduction band is pushed down, more electrons are injected from the source, and $n_s(0)$ is maintained approximately at the value given by eq. (1). This barrier lowering effect is seen in Fig. 3, and a plot of $\langle n_S(0)\rangle$ *vs.* V_{DS} confirms that in a "well-tempered MOSFET," which is designed to electrostatically isolate the drain from the source [10], MOS electrostatics maintain the inversion layer charge at the beginning of the channel at an approximately constant value. (The same effect has also been observed in 2D Monte Carlo simulations [11].)

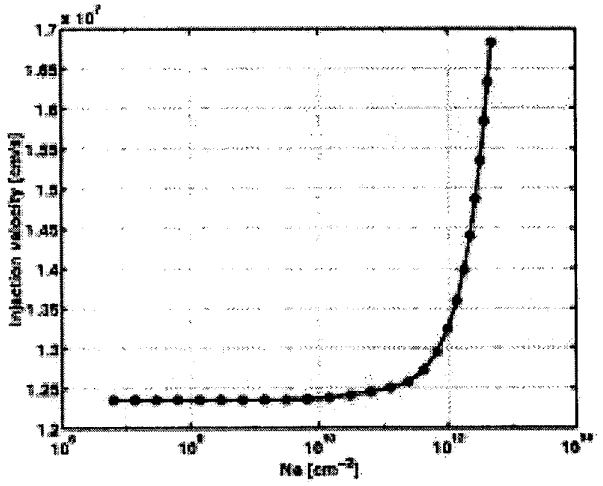

Fig. 4 The thermal injection velocity, υ_T vs. inversion layer density, n_S for the DG SOI MOSFET.

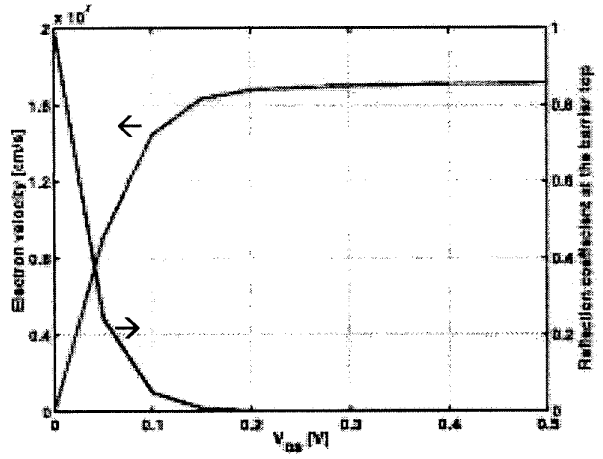

Fig. 5 The average velocity at the beginning of the channel vs. V_{DS} for the device of Fig. 3. For this gate $n_S \approx 5\times10^{12}\,\mathrm{cm}^{-2}$. Also shown is the ratio, r, of negative to positive flux, which is a measure of the anisotropy of the distribution (dashed line).

III. SCATTERING

In a ballistic MOSFET, positive velocity carriers come from the source and negative velocity carriers from the drain, but in a real MOSFET, scattering mixes the two streams. The result is that $\langle\upsilon(0)\rangle$ is less than υ_T. Assuming V_{DS} is greater than a few kT/q, then all negative velocity carriers arise from backscattering, and one can readily show

$$\langle\upsilon(0)\rangle = \left(\frac{1-r}{1+r}\right)\upsilon_T \tag{4}$$

where $0 < r < 1$ is the fraction of the injected carriers that backscatter and leave the channel. Well-designed MOSFETs currently operate with $r \approx 0.5$ [2], so from eq. (4) $\langle\upsilon(0)\rangle$ is about one-third of its limit, but devices with $r \approx 0.2$ have been recently reported [1]. Note that when eq. (4) is inserted into eq. (2), we get a result presented earlier [6, 12].

Since backscattering is what limits the drive-current of a MOSFET, so one should understand what controls r. The model calculations displayed in Fig. 6 address this issue. Thermal equilibrium carriers are injected into a constant-field region and tracked by Monte Carlo simulation (we used the model of [13] with an ionized impurity density of 3 x 10^{19} cm^{-3} which produces a low field mean-free-path of ≈9 nm). The simulations reveal that if carriers travel more than about kT/q down the potential, then even if they do scatter, they are unlikely to emerge from the channel and contribute to r. (Price noted a similar effect some time ago [14]). These observations are supported by the data plotted in Fig. 6, which shows the fraction backscattered vs. the maximum distance the carriers penetrated into the channel. The critical distance, ℓ, over which the first kT/q potential drop occurs is also noted. Note that when a high field is present, ~65% of the backscattered carriers only penetrated to the first kT/q of potential drop. These results show that r is most sensitive to scattering in the so-called kT layer, an effect that is closely analogous to the well-known Bethe condition for thermionic emission [15].

The backscattering coefficient for a field-free slab of length, L, is [16]

$$r = \frac{L}{L+\lambda_o}, \tag{5a}$$

where λ_o is the near-equilibrium mean-free-path. When an electric field is present, eq. (5a) must be modified to [6]

$$r = \frac{\ell}{\ell+\lambda_o} \tag{5b}$$

assuming that $\ell < L$. Note that we use the near-equilibrium mean-free-path because the scattering that contributes to r occurs in the initial kT layer, before carriers have been significantly heated. Equation (5b) generally agrees well with Monte Carlo simulations, which confirms that the near-equilibrium mobility is an important physical parameter because it controls backscattering in the critical portion of the channel.)

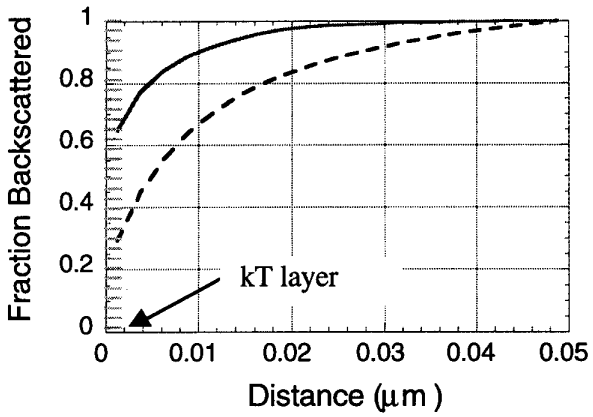

Fig. 6 The fraction backscattered vs. distance penetrated into the channel. The channel is 50nm long with a 0V (dashed line) or 1V applied across it.

Another way to view this problem is in terms of diffusion across the current limiting region. As suggested by Fig. 2, carriers must diffuse across the initial low-field part of the channel before they are collected by the high-field portion of the channel, which is a nearly perfect absorber for those carriers that reach it. This is the classic problem of diffusion across a thin base [17]. It's clear that in the limit that the region is short, or that scattering is weak, the maximum diffusion velocity is the thermal velocity, υ_T.

4. QUANTUM TRANSPORT

Quantum confinement effects on the threshold voltage and gate capacitance and quantum mechanical tunneling currents are now rather well understood, and techniques to treat them in simulations are available [18]. In nanoscale MOSFETs, however, the channel length is approaching dimensions that are comparable to the thickness of an inverion layer, so the role of quantum transport along the channel needs to be examined [19].

Figure 7 compare the ballistic *I-V* characteristics of the 10 nm DG SOI MOSFET as computed by classical and quantum approaches. (The quantum mechanical simulations use a Green's function method [20].) What is most surprising is how small the differences are, even at 10nm channel lengths. One important effect is the small increase in subthreshold swing (from 66 to 70 mV/dec) which is due to quantum mechanical tunneling through the source to channel barrier. The fundamental scaling limit of this device, which occur when carriers tunneling directly from source to drain, is apparently less than 10nm.

There is, of course, much more to consider with regard to transport at the quantum scale, but these simple calculations show that the MOSFET is essentially a classical device and that quantum interference effects are not expected to be strong.

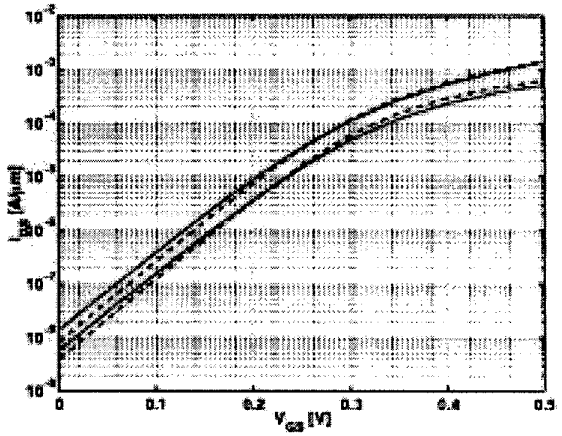

Fig. 7 The $\log(I_D)$ vs. V_{GS} characteristics of the 10nm DG SOI MOSFET as computed by a classical (dashed) and quantum mechnical (solid) simulation for V_D = 0.05 and 0.5V.

5. DEVICE TCAD

A critical issue now is to examine the validity of TCAD tools developed in the past for relatively large devices to the devices with nanoscale channel lengths that are now being developed. Figure 8 compares some drift-diffusion simulations to ballistic simulations. As discussed in the previous section, classical and quantum transport models give very similar predictions. The conventional drift-diffusion simulation predicts rather low current because the velocity within the channel is unphysically clamped at υ_{sat} as shown in Fig. 9.

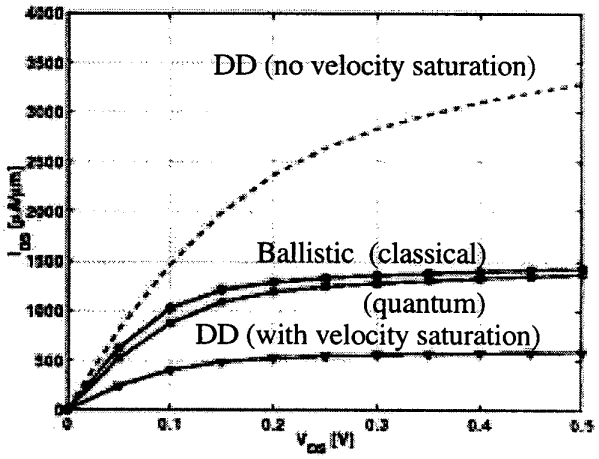

Fig. 8 I_D vs. V_D for V_G = 0.5V for the 10 nm DG-SOI MOSFET as computed by classical ballistic, quantum ballistic, drift-diffusion, and drift-diffusion with a constant mobility.

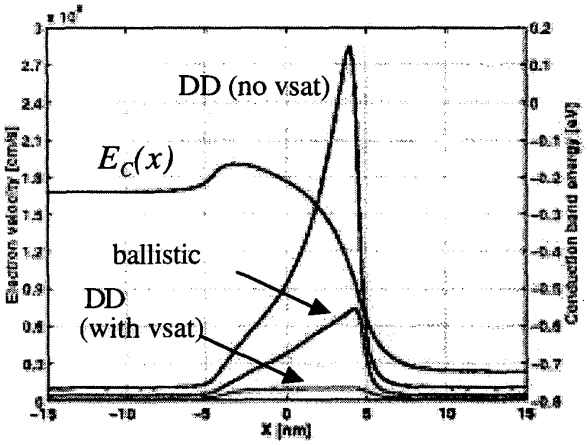

Fig. 9 Average velocity vs. position at $V_G = V_D = 0.5V$ for the 10 nm DG SOI-MOSFET. For reference, we also show the conduction band profile; the beginning of the channel is at the top of the barrier.

For comparision, we also show the results of a DD simulation with a constant mobility, so that velocity saturation does not occur. As shown in Fig. 8, the computed drain current exceeds the ballistic limit. It may be thought that this result is due to the unphysically high velocites near the drain, which are shown in Fig. 9. As Fig. 9 shows, however, the real problem is that the velocity at the beginning of the channel exceeds the thermal limit, v_T. Recall that carriers diffuse across the low-field region at the beginning of the channel. In the conventional drift-diffusion equation, there is no mechanism to limit the diffusion velocity to the thermal velocity. The limitation arises from the convective part of the kinetic energy, which is ignored in most transport models [21].

A proper treatment of carrier transport in nanoscale MOSFETs must allow strong velocity overshoot within the channel, but it must not allow the velocity to exceed the thermal limit at the beginning of the channel. Energy transport models were developed to treat hot carrier effects and velocity overshoot, but they also ignore the convective part of the kinetic energy. As a result, energy transport models don't enforce the thermal velocity limit at the beginning of the channel, and they generally over-predict the on-current of a nanoscale MOSFET [8]. The development of a transport model that works from the diffusive to ballistic regimes is a major challenge for the modeling and simulation community [8].

VI. SUMMARY

We presented a simple view of the essential physics of carrier transport in nanoscale MOS transistors and showed that in spite of the complex non-local transport that occurs in such devices, the physics that determines the steady-state current can be simply explained. With this conceptual view, the upper limit performance of a device is readily established, and the limitations of existing TCAD transport models are clearly illustrated. This view should prove useful in interpreting detailed simulations and in guiding the development of nanoscale transistors and TCAD models for them.

Acknowledgements —this work was supported by the Semiconductor Research Corporation.

REFERENCES

[1] G. Timp, A. Agarwal, F.H. Baumann, T. Boone, M. Buonanno, R. Cirelli, V. Donnelly, M. Foad, D. Grant, M. Green, H. Grossmann, S. Hillenius, J. Jackson, D. Jacobson, R. Kleiman, A. Kornblit, F. Klemens, J.T.C. Lee, W. Mansfield, S. Moccio, A. Murrell, M.O.'Malley, J. Rosamilia, J. Sapjeta, P. Silverman, T. Sorsch, W.W. Tai, D. Tennant, H. Vuong, B. Weir, "Low leakage, ultra-thin gate oxides for extremely high performance sub-100 nm nMOSFETs," *Conf. Record, International Electron Dev. Meeting*, Washington, D.C., Dec. 1997

[2] F. Assad, Z. Ren, P. Bendix, and M.S. Lundstrom., "Performance Limits of Si MOSFET's," IEDM Tech. Digest, p. 547, Dec. 1999.

[3] M.R. Pinto, E. Sangiorgi, and J. Bude, "Silicon transconductance scaling in the overshoot regime," *IEEE Electron Dev. Lett.*, **14**, pp. 375-378, 1993.

[4] S.E. Laux, and M.V. Fischetti, "Monte Carlo simulation of submicrometer Si n-MOSFET's at 77 and 300 K," *IEEE Electron Dev. Lett.*, **9**, pp. 467-469, 1988.

[5] C. Jungemann, S. Keith, M. Bartels, and B. Meinerzhagen, "Efficient full-band Monte Carlo simulation of silicon devices," *IEICE Trans. on Electronics*, **E82**, pp. 870-879, 1999.

[6] M.S. Lundstrom, "Elementary scattering theory of the MOSFET," *IEEE Electron Dev. Lett.*, **18**, pp. 361-363, 1997.

[7] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M.S. Lundstrom, "On the performance limits for Si MOSFET's: A theoretical study," *IEEE Trans. Electron Dev.*, **47**, 232-240, 2000.

[8] K. Banoo, J.-H. Rhew, M. Lundstrom, C.-W. Shu, and J.W. Jerome, "Simulating Quasi-Ballistic Transport in Si Nanotransistors," presented at the 7th International Workshop on Computational Electronics, Univ. of Glasgow, Glasgow, UK, May 22-25, 2000.

[9] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor, *J. Appl. Phys.*, **76**, pp. 4879-4890, 1994.

[10] H. Hu, J.B. Jarvis, L.T. Su, and D.A. Antoniadis, "A study of deep-submicron MOSFET scaling based on experiment and simulation," *IEEE Trans. Electron Dev.*, **42**, pp. 669-677, 1995.

[11] Jeff Bude, personal communitcation, Dec., 1999.

[12] S. Datta, F. Assad, and M.S. Lundstrom, "The Si MOSFET from a transmission viewpoint," *Superlattices and Microstructures*, **23**, pp. 771-780, 1998.

[13] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Reviews of Modern Physics*, **55**, 645, 1983.

[14] P.J. Price, "Monte Carlo calculation of electron transport in solids," *Semiconductors and Semimetals*, **14**, pp. 249-334, 1979.

[15] F. Berz, "The Bethe condition for thermionic emission near an absorbing boundary," *Solid-State Electron.*, **28**, 1007-1013, 1985.

[16] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, Cambridge, UK, 1997.

[17] F. Berz, "Diffusion near an absorbing boundary," *Solid-State Electron.*, **15**, 1245-1255, 1974.

[18] Yuan Taur and Tak Ning, *Fundamentals of VLSI Devices*, Cambridge University Press, Cambridge, UK, 1998.

[19] Y. Naveh and K.K. Likharev, "Modeling of 10-nm-Scale Ballistic MOSFET's," *IEEE Electron Dev. Lett.*, **21**, 242-244, 2000.

[20] R.K. Lake and S. Datta, "The Non-equilibrium Green's function method applied to double barrier resonant tunneling diodes, *Phys. Rev. B*, **45**, 6670-6685, 1992.

[21] G. Baccarani and M.R. Wordemann,"An Investigation of Steady-State Velocity Overshoot in Silicon." *Solid-State Electron.*, **28**, 407-416, 1985.