RC 20130 (07/14/95)
Engineering and Technology

IBM Research Report

Transport Models for Advanced Device Simulation – Truth or Consequences?

S.E. Laux and M.V. Fischetti

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Transport Models for Advanced Device Simulation – Truth or Consequences?

S.E. Laux and M.V. Fischetti
IBM Research Division; T.J. Watson Research Center
P.O Box 218; Yorktown Heights, NY 10598 USA

Abstract

An overview of three methods currently employed in advanced semiconductor device simulation is given, together with results for a model n^+-n-n^+ structure. The Monte Carlo, energy transport and spherical harmonic expansion methods are discussed from the perspective of a Monte Carlo enthusiast.

1 Introduction

How rigorous a description of carrier transport is required for semiconductor device modeling? This question is the subject of this paper, yet the question defies an absolute answer: the rigor of the model employed should be matched to the questions being asked. The trends in silicon device technology are towards smaller devices, operated at lower voltages. The usual arguments made concerning the former are that smaller dimensions imply fewer numbers of scattering events occur within the device active region, *i.e.*, the region responsible for active, rather than passive, character. As the active region shrinks, the applicability of any transport theory based upon averages over carriers having undergone a "large" number of scattering events becomes questionable. Just such an assumption underlies the present-day workhorse of semiconductor device modeling: the drift-diffusion (DD) transport model[1–4]. This paper discusses the most promising of the many alternative transport theories have been developed to address this well-known shortcoming of the DD theory. Each new theory is vying to replace DD as the new industry workhorse; computer implementations of many of these alternative theories are becoming available for use by device engineers. Throughout this paper, consistent with this practical bent, the semiconductor Si is implicit.

This paper discusses only theories derived from the Boltzmann transport equation (BTE)[5–7]

$$\frac{\partial f}{\partial t} + \nabla_{\vec{k}} f \cdot \frac{d\vec{k}}{dt} + \nabla_{\vec{r}} f \cdot \frac{d\vec{r}}{dt} = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \tag{1}$$

where $f(\vec{r}, \vec{k}, t)$ is the distribution function defined so that $f(\vec{r}, \vec{k}, t)\, d\vec{r}\, d\vec{k}/4\pi^3$ is the number of electrons at time t in the phase space volume $d\vec{r}\, d\vec{k}$ about the position $(\vec{r}, \vec{k})$. Theories founded primarily upon the physics of quantum transport[8] are relegated to the (distant) future so far as device engineers are concerned[1]. Such a "semi-classical" description utilizes quantum physics to describe scattering rates and band structure of the semiconductor, but otherwise, the BTE is a classical theory: carriers have simultaneously-known positions and momenta. The most promising theories are: approximate solution of the BTE via augmented drift-diffusion[10, 11], energy-transport and hydrodynamic formulations[12–16]; or solution of the BTE via cellular automata[17–19], scattering matrix[20–22], series expansion[23–26] or Monte Carlo[27–34] techniques. It is impossible to describe each method in detail; rather, this paper will be limited to two parts. First, three approaches – energy transport (ET), Monte Carlo (MC) and spherical harmonic expansion (SHE) – will be contrasted in the context of a model problem (an n^+-n-n^+ structure). Secondly, a critical assessment of each of the aforementioned approaches will be given. From both parts, the conclusion is the same: when interpolating between endpoints established from experimental data, most any model will do, but when extrapolating to new situations, or where physics issues are at stake, a Monte Carlo formulation based upon a complete representation of the semiconductor band structure is the approach of choice. Any simpler approach will mislead as often as it will guide, with uncertain consequence[2].

2 ET, MC and SHE Approaches

2.1 Energy Transport Approach

The ET models usually trace their lineage to Stratton[35]; the similar, but more complicated HD models trace their lineage to Bløtekjær[36]. In either case, the general approach to derive the theory is the same: multiply the BTE by powers of the momentum and integrate over momentum space. This accounts for ET and HD models being labeled "moment methods", as they both are derived from taking momentum moments of the BTE[5–7]. The zeroth moment (*i.e.*, multiplying by unity) yields a equation which expresses conservation of particle number. The first moment (*i.e.*, multiplying by momentum) yields an equation expressing conservation of momentum. The second moment yields an equation expressing conservation of energy. The third moment yields an equation expressing conservation of energy flux. This scheme only rigorously solves the BTE if all moments are accounted for; however, this is not done in order to render the calculation tractable. Rather, the hierarchy is terminated, almost always including only the zeroth, first and second moment equations. This is the first major assumption in these methods. In addition, this termination of the infinite hierarchy is a significant aspect of the theory, for the n^{th} moment equation contains

[1] Small devices and sublinearly-scaled biases imply large internal fields, which imply energetic carriers, which imply small de Broglie wavelengths. For realistic situations, the de Broglie wavelength is far smaller than the device active region size, which permits a semi-classical, rather than a quantum description, to be used[9].

[2] The reader should beware: having spent the last eight years working on full band Monte Carlo device simulation, we are not about to hide our bias on this subject!

explicitly the n+1th power of momentum, *i.e.*, the variable governed by the n+1th moment equation! How best to represent this unknown and unwanted dependence on the n+1th power of momentum in the nth equation leads to an assumption dubbed "closure". How best to close the moment equations is controversial and under active discussion[37–40], and represents the second major assumption in these methods. Finally, the scattering terms present in the BTE must be made commensurate with the variables on hand in the moment equations; this is done by adopting the "relaxation time approximation"[5–7, 41] and represents a third major assumption in these methods.

ET and HD models depart company at this point: HD models solve the resultant equations, retaining a carrier energy that depends both on average velocity (drift, or convective term $= \frac{1}{2}m^*v^2$) and deviations about the average velocity (thermal term $= \frac{3}{2}k_BT$). The ET models generally neglect the convective term, which can be a reasonable assumption depending on the device being modeled.

The ET model exercised in this paper is a very workable and mature implementation available from Integrated Systems Engineering AG[42]. Spherically symmetric, parabolic energy bands are assumed in deriving the ET equations to follow. In addition to the usual Poisson equation, this code solves the following partial differential equations for electron transport: (continuity of electron number, from the zeroth moment)

$$\nabla \cdot \vec{J}_n = qR + q\,\frac{\partial n}{\partial t}\ , \tag{2}$$

(conservation of electron energy, from the second moment)

$$\frac{\partial W_n}{\partial t} + \nabla \cdot \vec{S}_n = \vec{J}_n \cdot \vec{F} - R\,\frac{W_n}{n} - \frac{W_n - W_{n0}}{\tau_w}\ , \tag{3}$$

with the following constituent relations: the current density (obtained from the first moment, with some additional approximations)

$$\vec{J}_n = \mu_n(qn\vec{F} + k_BT_n\nabla n + k_Bn\nabla T_n)\ , \tag{4}$$

the energy flux (obtained from the third moment, with some additional approximation)

$$\vec{S}_n = -\frac{5}{2}r_J\frac{k_BT_n}{q}\vec{J}_n - r_T\kappa_n\nabla T_n\ , \tag{5}$$

and the Wiedemann-Franz Law for an electron gas[43]

$$\kappa_n = \frac{5r_T}{2}\frac{{k_B}^2}{q^2}q\mu_n nT_n\ , \tag{6}$$

where $\vec{J}_n$ (A/cm^2) is the electron current density, R (cm^{-3} sec^{-1}) is the electron recombination rate, W_n (J/cm^3) is the electron energy density, W_{n0} is the equilibrium electron energy density, $\vec{S}_n$ (W/cm^2) is the electron energy flux, $\vec{F}$ (V/cm) is the electric field, n (cm^{-3}) is the electron density, τ_w (sec) is the energy relaxation time, μ_n (cm^2/(V sec)) is the electron mobility, q (C) is the magnitude of the electronic charge, k_B (J/K) is Boltzmann's constant, T_n (K) is the electron temperature, and $r_J = r_T$ is the ratio (τ_s/τ_m) of the energy flux to momentum relaxation times. Theoretically, $r_J = r_T = 1 + \frac{2}{5}c$, where c is the exponent of the energy dependence of τ_m, *i.e.*, $\tau_m \propto (W_n/n)^c$. For the ET simulations in this paper, μ_n (and hence τ_m) is similar to the prescription of Ref. [44] and a constant τ_w is used, as described in Ref. [42]. Use of a constant τ_w is somewhat simplified[40].

2.2 Monte Carlo Approach

Monte Carlo methods can be used to obtain a (stochastic) solution to the BTE[28, 31]. There is *no* approximation to the form of the BTE employed by the method. Supplementing Eq. (1) are the constituent relations

$$\frac{d\vec{r}}{dt} = \frac{1}{\hbar}\,\nabla_{\vec{k}}E(\vec{k})\ , \tag{7}$$

$$\frac{d\vec{k}}{dt} = -\frac{q}{\hbar}\,\vec{F}(\vec{r})\ , \tag{8}$$

where $E(\vec{k})$ is the energy versus wavevector (band structure) relation for the semiconductor, $\vec{F}(\vec{r})$ is the electric field and $\hbar$ is the reduced Planck's constant. The approximations which discriminate between various implementations of MC transport in the literature involve what band structure $E(\vec{k})$ and collision terms $(\partial f/\partial t)_{coll}$ are utilized. This statement cannot be overemphasized: *all MC programs are not created equal*. These are the two places the "physics" enters the MC transport model, *i.e.*, the models for bands and the models for scattering rates. There are no mobility models *per se* – only transport physics that imply a certain mobility.

The MC device simulator DAMOCLES is used to obtain the MC results presented in this paper[30, 33, 34]. This program is the product of over 16 man-years of effort. DAMOCLES is used in two different ways: as an implementation of a "simple" parabolic band structure model, and as a "full band" MC implementation. The parabolic (dependence on wavevector) band structure maintains ellipsoidal equi-energy surfaces:

$$E(\vec{k}) = \frac{\hbar^2}{2}\sum_{i=1}^{6}\left[\frac{\left(k_x - k_{x0}^{(i)}\right)^2}{m_x^{(i)}} + \frac{\left(k_y - k_{y0}^{(i)}\right)^2}{m_y^{(i)}} + \frac{\left(k_z - k_{z0}^{(i)}\right)^2}{m_z^{(i)}}\right]\ , \tag{9}$$

where $\vec{k} = (k_x, k_y, k_z)$ and the ith ellipsoidal energy valley is centered about $(k_{x0}^{(i)}, k_{y0}^{(i)}, k_{z0}^{(i)})$ with effective mass $(m_x^{(i)}, m_y^{(i)}, m_z^{(i)})$. The $^{(i)}$ superscript refers to the six ellipsoidal energy valleys: the two valleys along the k_x-axis are centered at $\frac{2\pi}{a} \times (\pm 0.85, 0, 0)$ with masses $m_0 \times (0.92, 0.19, 0.19)$, the two along the k_y-axis are centered at $\frac{2\pi}{a} \times (0, \pm 0.85, 0)$ with masses $m_0 \times (0.19, 0.92, 0.19)$ and the two along the k_z-axis are centered at $\frac{2\pi}{a} \times (0, 0, \pm 0.85)$ with masses $m_0 \times (0.19, 0.19, 0.92)$. Here, a is the lattice constant of Si and m_0 is the free electron mass. In contrast, the full band model implements the first five conduction bands of Si obtained from empirical pseudo-potential band structure calculations as a look-up table within the program, *i.e.*, without any additional approximations to the form of the bands.

The scattering mechanisms $(\partial f/\partial t)_{coll}$ employed in DAMOCLES are all of the relevant processes, including phonon (optical and acoustic), impurity, carrier-carrier (electron-electron, hole-hole and electron-hole), alloy and impact ionization scattering. Each process is implemented using the best available physics within the constraint that the final program must execute in a tolerable amount of computer time.

2.3 Spherical Harmonic Expansion Approach

An approximate numerical solution to the BTE can be obtained by expressing the distribution function $f(\vec{r}, \vec{k}, t)$ as a series expansion, substituting this expression into Eq. (1), and solving for the expansion coefficients. Once $f(\vec{r}, \vec{k}, t)$ is known, macroscopic quantities like carrier density, energy, current, etc., follow from integrations over $\vec{k}$ space. A popular approach has been the SHE approach, being first invoked prior to large-scale computer modeling, but lately, experiencing a rediscovery of sorts. The spherical harmonics are $Y_l^m(\theta, \phi)$, $l = 0, 1, \ldots$, and $m = -l, -l+1, \ldots, 0, \ldots, l-1, l$, and are useful in expanding the distribution function $f(\vec{r}, \vec{k}, t)$ in polar phase-space coordinates $\vec{k} = (k, \theta, \phi)$, where $k = |\vec{k}|$, $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$, according to[45]

$$f(\vec{r}, k, \theta, \phi, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} b_{lm}(\vec{r}, k, t) \, Y_l^m(\theta, \phi) \,. \qquad (10)$$

Rarely is the rigor and generality of Eq. (10) preserved in semiconductor modeling; rather, two important approximations are invoked. First, the expansion is truncated at some finite value of l (usually one!) and secondly, a simplification in angular dependence is anticipated as follows. Assuming a spherically symmetric band structure about $\vec{k} = \vec{0}$, the distribution function f is also spherically symmetric about $\vec{k} = \vec{0}$ in equilibrium, and is expected in nonequilibrium to rigidly shift away from $\vec{k} = \vec{0}$ in a direction colinear with the applied field $\vec{F}$ (to first order). Therefore, by chosing the angle θ according to $\vec{k} \cdot \vec{F} = |\vec{k}| \, |\vec{F}| \cos\theta$, the distribution function is expected to have no dependence on the azimuthal angle ϕ[3]. This choice of θ is tantamount to expanding in wavevector polar coordinates referenced to the (local) field direction. If there is no ϕ dependence, the summation over m in Eq. (10) must be restricted to $m = 0$ only; with this simplification, Eq. (10) becomes

$$f(\vec{r}, k, \theta, t) = \sum_{l=0}^{\infty} b_l(\vec{r}, k, t) \, P_l(\cos\theta) \,, \qquad (11)$$

where $P_l(x)$ is the l^{th} Legendre polynomial[45]. Since $P_0(x) = 1$ and $P_1(x) = x$, the expansion up to $l = 1$ is simply $f = b_0 + b_1 \cos\theta$, a form which has seen usage for many years[41].

The approach in this paper will not be to exercise a device modeling program based upon Eq. (11), but rather, distribution functions obtained from MC modeling will be decomposed into their spherical harmonic expansions, in order to test the theory's truncation assumption.

3 The n^+-n-n^+ Structure

The one-dimensional n^+-n-n^+ "device" (actually more like a glorified resistor!) has often been simulated when discussing new device modeling techniques or transport theories. This stature is a bit undeserved, as this two-terminal structure is not in the least interesting to device designers. Nevertheless, the internal potential present bears resemblance to the potential present in metal-oxide-semiconductor field-effect transistor (MOSFET) channels or bipolar junction transistor (BJT) active regions. Fig. 1 shows the conduction band-edge potential energy versus position for a nominal bias and dimension, and serves to define the n^+-n-n^+ structure discussed in this paper. Electrons flow from cathode (source) to anode (drain) as a positive bias is applied at the anode. In this paper, the cathode and anode are separated by a distance L (= 0.2 μm in Fig. 1), with the cathode-channel doping transition located at the origin. The cathode and anode are doped $N_D = 10^{20}$ cm^{-3} and are 0.1 μm in length; the channel is doped $N_D = 10^{16}$ cm^{-3}. The cathode is grounded and a bias V_A (= 0.8 V in Fig. 1) is applied at the anode. Values $0.05 \leq L \leq 0.80\,\mu$m and $0.20 \leq V_A \leq 2.70$ V are considered here, consistent with future generations of channel lengths and supply voltages. Note the n^+ dopings used here are quite high, as is often the case in realistic semiconductor devices. The potential energy of Fig. 1 was obtained from an ET simulation.

[3] In 1D $\vec{r}$ space, f is independent of ϕ, so this is rigorous for the simple calculations here.

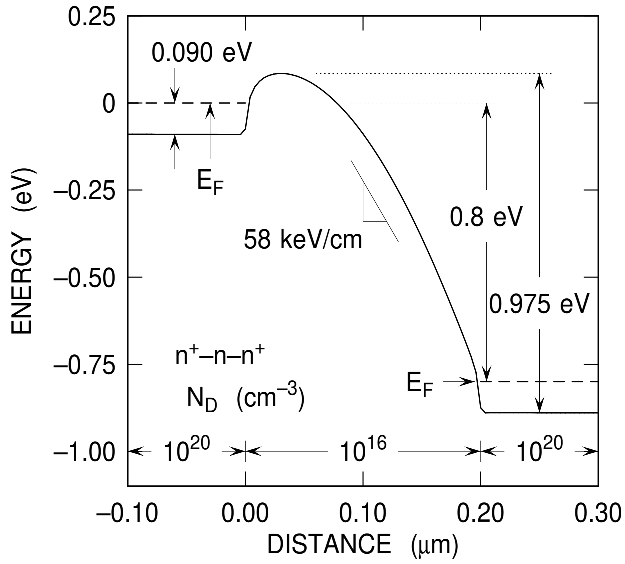

Fig. 1: Conduction band-edge potential energy versus position for a typical n^+-n-n^+ structure with $L = 0.2$ μm and $V_A = 0.8$ V. Because the cathode and anode are degenerately doped, the Fermi energies E_F reside 90 mV into the conduction band. While the Fermi levels are separated by the applied bias V_A, note that the peak of the potential barrier is actually 0.975 V above the band-edge in the anode.

4 MC Results

A Monte Carlo approach using a simple ellipsoidal, parabolic energy band model (Eq. (9)) and consistent scattering rates[28] is applied to the n^+-n-n^+ structure with $L = 0.20$ μm and $V_A = 0.8$ V. The conduction band-edge energy used is extracted from an ET simulation of the same structure, in order to reduce differences between MC and ET in anticipation of the comparisons discussed in the next section. Fig. 2 shows the distribution of electron energy at five different locations within the structure, as well as the conduction band-edge energy. The distributions labeled 'A' and 'E' correspond to the energy sampled over the center 0.05 μm of the cathode and anode, respectively. In the cathode ('A'), the anticipated Maxwellian thermal distribution is seen: $f(E) \propto (E - E_c)^{1/2} \exp(-(E - E_c)/k_B T_n)$, where E_c is the local conduction band-edge energy, $T_n = 300$ K because this is a thermal distribution and the term $(E - E_c)^{1/2}$ originates from the density-of-states in the conduction band. Note Fig. 2 has a dotted line corresponding to a 300 K temperature on each distribution. In the anode ('E'), the distribution is the superposition, in amounts proportional to their relative densities, of

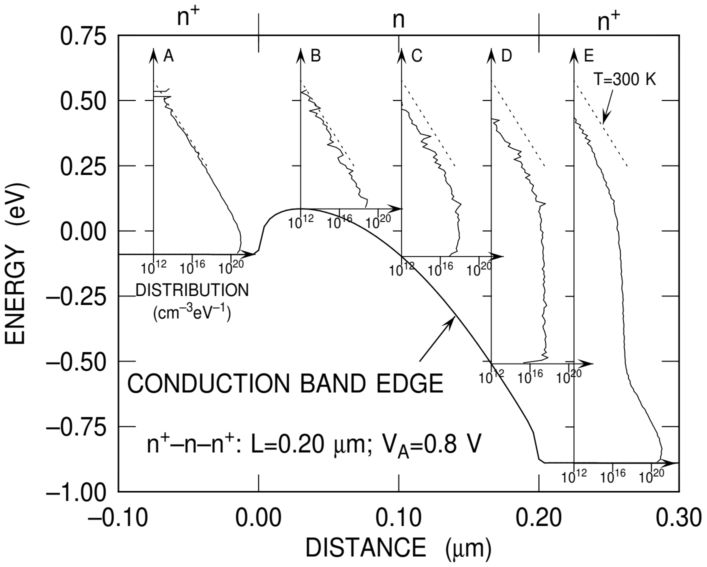

Fig. 2: The conduction band energy of a $L = 0.2$ μm n^+-n-n^+ structure with $V_A = 0.8$ V, together with electron energy distributions at five positions within the structure. Distributions marked 'A' and 'E' were sampled in the central 0.05 μm of the cathode and anode, respectively. Distributions 'B', 'C' and 'D' were sampled in a 0.004 μm interval centered at the position where the distribution function touches the band edge curve. An ellipsoidal, parabolic band structure model was used in this MC calculation. In each distribution, a dotted curve shows the asymptote $f \propto \exp(-E/k_B T_n)$, where $T_n = 300$ K.

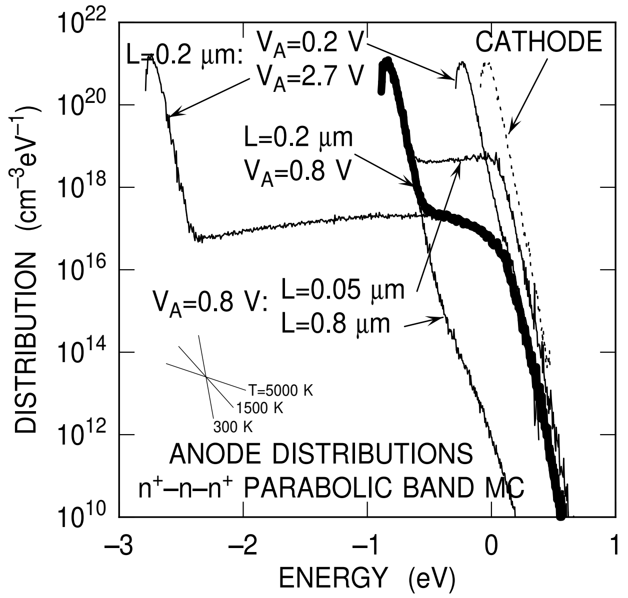

Fig. 3: Five distributions of electron energy in the anode of an n^+-n-n^+ structure. The nominal case ($L = 0.2$ μm, $V_A = 0.8$ V) is shown as a bold line. Four variations in size and bias are also considered: $L = 0.05$ or 0.8 μm ($V_A = 0.8$ V) and $V_A = 0.2$ or 2.7 V ($L = 0.2$ μm). An ellipsoidal, parabolic band structure model was used in these calculations. A generic cathode distribution (they are all the same) is shown for comparison. The energy scale is defined as in Fig. 2. Three slopes corresponding to the asymptote $f \propto \exp(-E/k_B T_n)$, $T_n = 300$, 1500 and 5000 K, are shown for reference.

two populations: a high density, thermal population of carriers in the anode, and a lower density, "hot" population which has traversed from cathode to anode. The manner in which this latter population traversed the structure from cathode to anode is shown by the distributions 'B', 'C' and 'D': these distributions sample the electron energy in a 0.004 μm interval centered at the location where each energy distribution "touches" the band edge in Fig. 2. In 'B', the distribution is essentially that of 'A', except for a skimming away of the lowest energy carriers due to the energy barrier. At 'C' and 'D', the distribution is increasingly more energetic, becoming elongated as some carriers scatter and move towards the local band-edge, while others scatter less, or not at all, or even absorb a phonon(s), and hence remain at higher energy. Note in every case, at energies above the peak of the potential barrier energy, *all distributions tend asymptotically towards a logarithmic slope of* $(k_B T_n)^{-1}$ *with* $T_n = 300$ *K*, as shown by a dotted line on each distribution. This has been noted previously[46, 47].

Fig. 3 presents five different anode distributions, analogous to distribution 'E' in Fig. 2, for a nominal structure $L = 0.2$ μm $V_A = 0.8$ V and four variations: $L = 0.2$ μm with $V_A = 0.2$ or 2.7 V, and $V_A = 0.8$ V with $L = 0.05$ or 0.8 μm. A generic cathode distribution, analogous to distribution 'A' in Fig. 2 is also shown (all five cases have the same cathode distribution). The nominal result is shown as a bold line, and corresponds to 'E' in Fig. 2. For a constant size $L = 0.2$ μm, as V_A is reduced to 0.2 V the anode distribution degenerates to a thermal result shifted 0.2 eV lower in energy than in the cathode. As V_A is increased to 2.7 V the anode distribution clearly divides into three regions: a high density, thermal population ($-2.79 \leq E \lesssim -2.4$ eV), a low density, energetic population ($-2.4 \lesssim E \lesssim 0$ eV), and an asymptotic thermal tail ($E \gtrsim 0$ eV). For a constant applied bias $V_A = 0.8$ V, as L increase to 0.8 μm the field is reduced so that the anode distribution has only a hint of a low density, energetic population representing electrons traversing the structure. As L decreases to 0.05 μm the anode distribution is again a union of three distinct regions, although more compressed in energy compared to the $L = 0.2$ μm, $V_A = 2.7$ V case.

The preceding MC results were based upon a simple MC transport model. Fig. 4 quantifies the answer to two important questions: how do the anode distributions change as the transport model becomes more complex, and secondly, how well can MC results be described by a "Maxwellian" or "Displaced Maxwellian" form[4]?

Fig. 4 shows anode distributions for a $L = 0.05$ μm, $V_A = 0.8$ V obtained with an ellipsoidal, parabolic band structure transport model as well as the full band DAMOCLES model, both with and without the presence of short-range electron-electron scattering. The full band result shows a lower average energy (see Fig. 5), consistent with the large number of available states at lower energies in the full band structure. Otherwise, the overall shape of the distribution is quite similar. The full band result including electron-electron scattering shows additional structure due to the possibility of an energetic electron entering the anode and gaining significant energy via e-e scat-

[4] A displaced Maxwellian distribution is defined by $f_{DM}(\vec{k}) \propto \exp(-\hbar^2|\vec{k}-\vec{k_0}|^2/2m^* k_B T_n)$, which applies in nondegenerate situations only. Assuming parabolic bands and a displacement in the k_x direction, then $f_{DM}(E) \propto (k_B T_n/\sqrt{E_0})\exp(-(E+E_0)/k_B T_n)\sinh(2\sqrt{EE_0}/k_B T_n)$, with $E_0 = \hbar^2 {k_0}^2/2m^*$. Since the displacement is often negligible compared to the breadth of the distribution ($E_0 \ll k_B T_n$), a simple Maxwellian is often appropriate: $f_M(E) \propto E^{1/2}\exp(-E/k_B T_n)$[33].

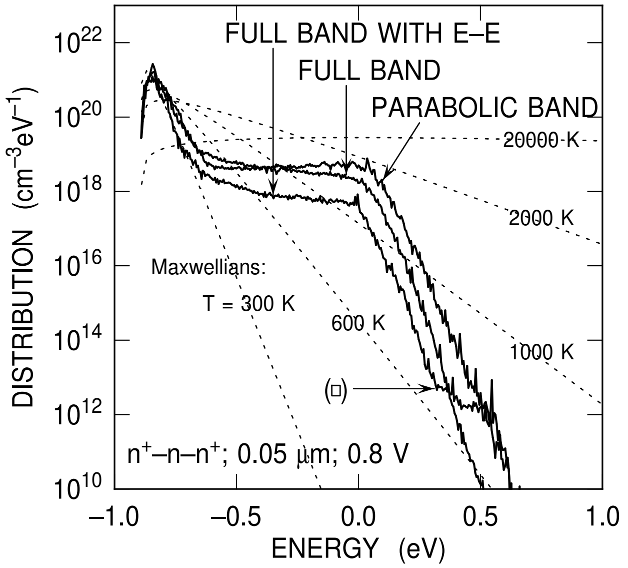

Fig. 4: Three electron energy distributions for an n^+-n-n^+ structure with L = 0.05 μm and V_A = 0.8 V, each based upon MC transport models of differing rigor. Six Maxwellian distributions at temperatures from 300 to 20000 K representing the same spatial density of electrons are shown as dotted curves. The MC distribution based upon a full band model including electron-electron (e-e) scattering shows an enhanced high energy tail, whose onset is indicated by (□).

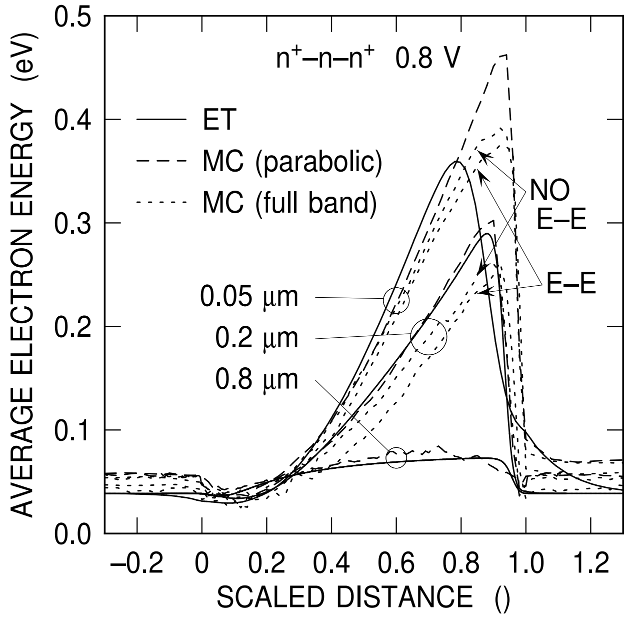

Fig. 5: Average electron energy versus position for a n^+-n-n^+ structures with L = 0.05, 0.2, 0.8 μm and V_A = 0.8 V, obtained from ET and parabolic band MC modeling. Cathode to anode distance is scaled to span from 0 to 1. For the smallest two structures, full band MC results are shown, both with and without electron-electron scattering.

tering. This is seen by the presence of an enhanced high energy tail whose onset is marked with a (□) in Fig. 4. Note that given the high electron density of the anode, a dynamic screening model[34] is required to obtain this result. In addition, e-e scattering depopulates the middle energy range $-0.6 \leq E \leq 0.2$ eV: the majority of these electrons move to lower energies (barely visible on the log scale of Fig. 4), while a few take up residence in the aforementioned enhanced high energy tail. Hence, e-e scattering potentially introduces another structure in the distribution function: an enhanced high energy tail. Were this distribution function obtained using a self-consistent potential solution which time resolved the potential fluctuations within the anode, this distribution would change yet again: all of the distinct slope changes would be somewhat smeared out, and the high energy tail would be further enhanced given that energetic electrons can now absorb energy from the oscillations of the "plasma" of low energy electrons in the anode. *Getting the precise high energy character of distribution functions correct in the anode is complicated, yet this is precisely where impact ionization would occur in a MOSFET whose channel potential is analogous to the n^+-n-n^+ potentials considered here.*

Fig. 4 also shows Maxwellian distributions for the same spatial density of carriers at five temperatures: T_n = 300, 600, 1000, 2000 and 20000 K. They bear no resemblance to the MC distributions; using a displaced Maxwellian does little to improve the fit. While using a superposition of *two* (displaced-)Maxwellians can better approximate the MC distributions, this still misses the 300 K slope obtained at energies above the peak of the potential barrier. Note that homogeneous MC results (simulations of electrons accelerated in a constant electric field) due bear great resemblance to the Maxwellians shown in Fig. 4. The explicit message is that distribution functions of devices can differ appreciably from those of homogeneous field simulation, because the latter is local in character, while in the former the distribution function depends sensitively on nonlocal electron behavior, *i.e.*, the manner in which electron behavior at different positions contributes to the final result at a given position.

5 ET vs. MC Results

Fig. 5 shows the average electron energy obtained from three ET simulations: L = 0.05, 0.2 and 0.8 μm at a fixed bias V_A = 0.8 V. The average energy is obtained from the ET model by multiplying the electron temperature T_n by $\frac{3}{2}k_B$; this assumes the drift energy $\frac{1}{2}m^*v^2$ is small by comparison, which is the case (especially near the anode). Therefore, the ET results of Fig. 5 are actually a lower bound for the true average energy. As expected, the energy rises as the size shrinks, given that the accelerating field grows larger. Also shown are MC simulation results. For these MC results, the conduction band-edge potential obtained from the respective ET simulation was utilized by the MC calculation, *i.e.*, MC simulation was based upon an electric field taken from an ET solution. This decision was made in order to reduce the number of variables between ET and MC simulation, for MC does not obtain exactly the same values of electric field in a fully self-consistent calculation. For a parabolic band structure model, results are shown at all dimensions as a dashed line. Agreement is excellent for L $\geq$ 0.2 μm, but this is by design: the values τ_w = 0.45 psec, r_J = 0.72 and r_T = 0.16 were selected to obtain this agreement at L = 0.2 μm. Since the ET model is based upon a parabolic band structure, agreement should then be outstanding for all dimensions. Yet, for the smallest dimension (L = 0.05 μm), ET underestimates the average energy. Note the ET model fails to predict correct average energies in the cathode: $\frac{3}{2}k_BT \approx 39$ meV at 300 K is *not* the

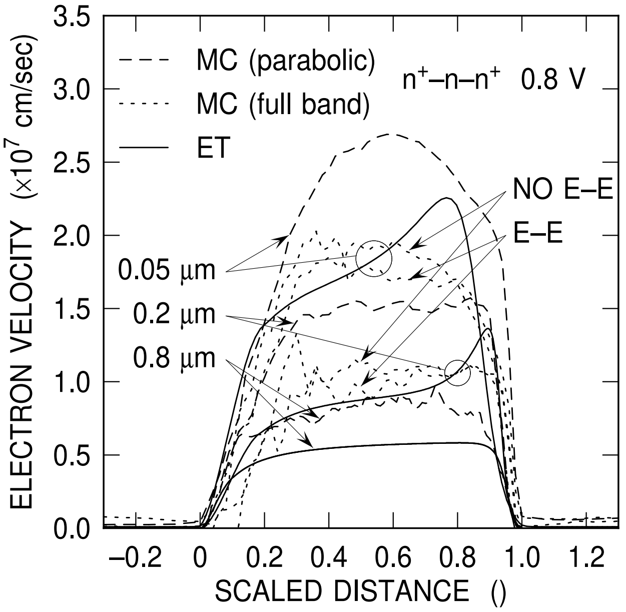

Fig. 6: Average electron velocity versus position for the same structures and transport models as shown in Fig. 5.

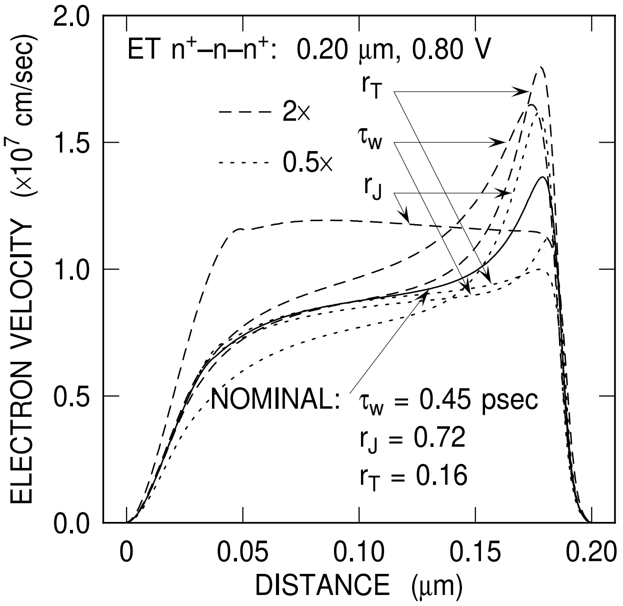

Fig. 7: Average electron velocity versus position for the 0.2 μm, 0.8 V n^+-n-n^+ structure using ET modeling for nominal τ_w, r_J and r_T parameters, as well as six perturbations. Each parameter is changed by either 2× (three dashed lines) or by 0.5× (three dotted lines) in turn, while the other two parameters remain unchanged.

correct average energy for $n = 10^{20}$ cm^{-3} concentrations, due to degeneracy. This error appears not to seriously effect the solution in the channel, however. Additional MC results are shown based upon a full band structure model, both with and without electron-electron scattering included, at the two smallest dimensions. The full band structure results in a lower average energy due to the significant differences between a parabolic bands and full bands away from the conduction band minima: there are many more opportunities for electrons to remain at lower energy and velocity away from the conduction band minima in the full band model. Electron-electron scattering does little to change the average energy, as expected.

Fig. 6 shows the average electron velocity for the same structures and conditions as in Fig. 5. Agreement here is less satisfying: parabolic band MC yields higher velocities than the ET model (although being based upon the same band structure); the lower velocities of the full band models are more similar to the ET velocities in magnitude, but not in shape. Electron-electron scattering makes a relatively larger difference in the MC results near the cathode than near the anode, consistent with the higher densities near the cathode. Note all models predict velocity overshoot, *i.e.*, velocities above $v_{sat} \approx 10^7$ cm/sec in Si at 300 K.

The MC-ET comparison summarized above (and many more simulations not shown) leads to a first conclusion: *ET can either yield correct average energies, or correct velocities, but not both.* That is, ET can be calibrated to yield average energies in agreement with parabolic band MC, at which point velocities will differ substantially, as was done here. If velocities are matched, energies will then differ substantially.

Furthermore, the results obtained from the ET model are quite sensitive to the model parameters. While this can be said for many modeling approaches, the problem is exacerbated for ET because there is little formal guidance for setting parameters: parameters are defined from the theory as integrals over the distribution function, with the result (i) since the distribution functions is unknown, the theory is of little assistance and (ii) the parameters are *functionals*, and not simply functions, of densities, fluxes, etc.[48, 7]. The net result is a kind of phenomenology surrounding the selection of parameters for ET[5]. Fig. 7 shows electron velocity versus position for the 0.2 μm, 0.8 V n^+-n-n^+ structure. Three parameters are varied: τ_w, r_J and r_T. These parameters were selected because they are some of the most "controversial" with respect to their recommended values in the literature. Besides the nominal result, six curves are shown for the cases when each of the three parameters is changed in turn by either 2× (three dashed curves) or 0.5× (three dotted curves). Changes in τ_w change the velocity near the anode ~20%, and the average energy near the anode ~30% (latter not shown). Such a change in energy (temperature) implies a significant change in hot electron population: *if* the distribution function is Maxwellian, the number of carriers above 1 eV change by 3.2× if the temperature increases from 2300 K to 3000 K, or by 34× for carriers above 3 eV. If impact ionization (onset of ~1 eV) or injection into an oxide (onset of ~3 eV) are of interest (and are not hot electrons the reason for doing this modeling in the first place?!), a 30% change in temperature must be seen as significant. Changes in r_J change the shape of the velocity *substantially*, as seen in Fig. 7. The energy changes by ~15% near the anode (not shown). Changes in r_T change the velocity peak in Fig. 7 substantially, but leave the energy virtually unchanged (not shown). Recent theory[49] suggests r_T should depend on the energy flux; this has the favorable effect of suppressing the unphysical velocity peak near the anode, a source of consternation for many years for ET modelers[14, 50].

A second conclusion, then, is *ET is sensitive to its parameters, especially with respect to the implied population of hot electrons, yet the theory is not always helpful in pinning down the necessary parameters.* Because parameters are functionals

[5] This is a continuation of a similar mentality surrounding the search for the ultimately correct mobility model for DD modeling.

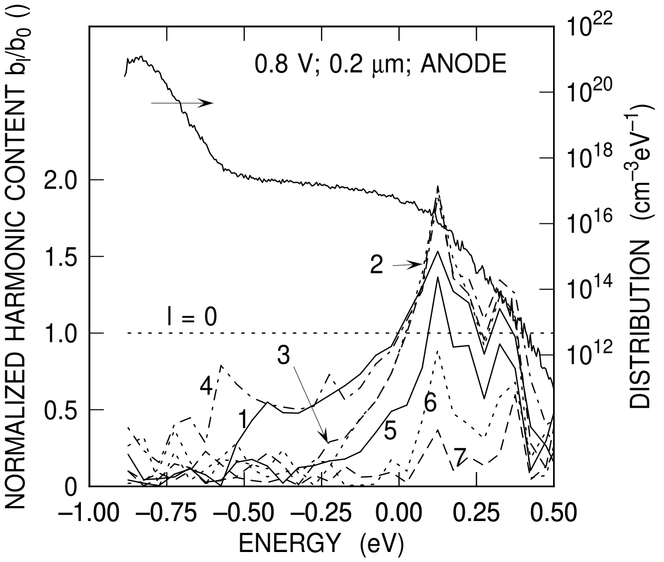

Fig. 8: Normalized harmonic content b_l/b_0 versus energy in the anode of an n^+-n-n^+ structure with $L = 0.2\ \mu$m and $V_A = 0.8$ V. The spherical harmonic index l from 0 to 7 is shown; since the distribution is mapped into a spherically symmetric phase space by a Herring-Vogt transformation, there is no ϕ dependence and the index m = 0 is implicit. The corresponding distribution in energy is shown as well.

involving integrals of the distribution function (which is nonlocal in character) it follows that spatially-*homogeneous* MC (whose distribution function is local in character) can *not* be used to evaluate ET parameters.

6 SHE vs. MC Results

In order to ascertain the validity of a SHE approach to electron transport, and especially the necessity of truncating the expansion to finite order, the ellipsoidal parabolic band MC transport model was exercised to obtain anode distributions in $\vec{k}$ space. By applying a Herring-Vogt transformation[51] to $\vec{k}$ space, the distributions can be made spherical in $\vec{k}^*$ space. These distributions in $\vec{k}^*$ space are then expanded according to Eq. 11, since $f(\vec{k}^*)$ is independent of ϕ. The results shown in Fig. 8 are typical of many cases examined. Fig. 8 shows the ratio b_l/b_0 as a function of energy for the anode distribution of a 0.2 μm n^+-n-n^+ structure with $V_A = 0.8$ V; the actual distribution in energy is shown as well. As indicated previously, often only the $l = 0$ and 1 harmonics in Eq. (11) are included. Admittedly, the harmonic content results of Fig. 8 suffer from statistical noise: gathering distributions in energy is challenging for MC; gathering distributions in energy *and* $\cos\theta$ compounds the challenge. Nevertheless, the results are considered essentially correct: the harmonic content is "low" for the high density, thermal electrons in the anode ($E < -0.6$ eV), but grows markedly as 0 eV is approached. Truncating the SHE approach at $l = 1$ is seen to be completely unjustified. To compound the problem, using the SHE requires *either* assuming spherical bands (in original $\vec{k}$, or transformed $\vec{k}^*$ space) to eliminate a ϕ dependence in the expansion, *or* using a realistic band structure and retaining many extra terms to resolve a ϕ dependence in the expansion. Thus, the method is intrinsically limited.

7 Discussion

While an n^+-n-n^+ structure is not a very sophisticated vehicle to exercise transport theories for advanced device simulation, it does have the advantage of providing a simple testbed where some plain truth can be demonstrated. Distributions in energy can be quite complex, especially at the anode of an n^+-n-n^+ structure (drain of a MOSFET): here hot and cold carriers mix, yet at sufficiently high energies, a 300 K slope is obtained. Carrier-carrier scattering enhanced the lowest and highest energy populations, at the expense of middle energies. Plasma oscillations smooth out these distinctive features somewhat. Nevertheless, these complexities can result in orders-of-magnitude differences in quantitative measures of hot electron population, *e.g.*, in the substrate current of a MOSFET[34, 52]. Where does this truth leave methods other that a full band MC approach?

ET methods (and their more sophisticated relatives HD methods) are hampered by simplified band structures, truncation of the moment expansion, closure relations, and intrinsic functional dependence of parameters on the (unknown) distribution function. Noteworthy progress in moment methods includes: inclusion of second- and fourth-order moments[53], new insights on closure[40], two-gas models[54], use of nonlocal MC to provide problem-dependent parameter values[55, 56], better band structure models[57, 58], and new expansions based upon perturbations from pure ballistic transport[59]. While these methods are evolutionary extensions of DD, their underpinnings are still disconcerting[60].

SHE methods seem promising if the expansion is performed to a sufficiently high order, but this has generally not been the case.

Finally, a few critical remarks on the other solution techniques mentioned at the outset of this paper are in order. Augmented DD is really intended as a "stop-gap" theory: a minimalist approach that can be retrofitted into an existing DD computer code with little effort. It is not really suitable for the modeling of complex devices. Cellular automata is promising, rigorous and capable. It may be positioned to compete favorably with MC; however, is it superior to MC? As a stochastic method, the cellular automata approach has similar statistic noise problems and computational appetite as MC. Merely demonstrating parity with MC will not gather much support for the cellular automata approach. Finally, the scattering matrix approach is a method held in high esteem by many. The method appears to have many advantages. The singular disadvantage appears to be the challenge of exploiting the method in multiple $\vec{r}$ space dimensions. In 1D, the method is impressive. In 2D, the method is uncharted. Whether the method maintains an advantage of computational speed over MC in 2D and 3D remains an unanswered question.

8 Conclusions

The MC, ET and SHE methods of modeling electron transport for advanced device simulation have been discussed. Where rigor is required, a full band MC approach like that of the DAMOCLES program remains the method of choice. Other methods may accomplish simpler tasks with less computational burden, but today MC remains the best tools to garner physical insight into the operation of practical, ultra-miniature semiconductor devices.

9 Acknowledgement

Conversations with Prof. M. Anile and Prof. P. Degond are gratefully acknowledged.

References

[1] W. Van Roosbroeck, *B.S.T.J.*, **29**, 560, 1950.
[2] A.H. Marshak and C.M. Van Vliet, *Proc. IEEE*, **72**, 148, 1984.
[3] S. Selberherr. *Analysis and Simulation of Semiconductor Devices.* Springer-Verlag, Vienna, 1984.
[4] W. Hänsch. *The drift-diffusion equation and its applications in MOSFET modeling.* Springer Verlag, Vienna, 1991.
[5] K. Huang. *Statistical Mechanics.* Wiley, New York, 1963.
[6] F. Reif. *Fundamentals of Statistical and Thermal Physics.* McGraw-Hill, New York, 1965.
[7] Karl Hess. *Advanced Theory of Semiconductor Devices.* Prentice Hall, Englewood Cliffs, NJ, 1988.
[8] W. Kohn and J.M. Luttinger, *Phys. Rev.*, **108**, 590, 1957.
[9] Karl Hess. Panel session on 'Device Modeling Below 0.1 Micron'. Third International Workshop on Computational Electronics, May 1994.
[10] K.K. Thornber, *IEEE Electron Device Lett.*, **3**, 69, 1982.
[11] M. Artaki, *Appl. Phys. Lett.*, **52**, 141, 1988.
[12] G. Baccarani and M.R. Wordeman, *Solid-State Electronics*, **28**, 407, 1985.
[13] A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan, and G. Baccarani, *IEEE Trans. Computer-Aided Design.*, **7**, 231, 1988.
[14] A. Gnudi, F. Odeh, and M. Rudan, *European Trans. Telcomm.*, **1**, 307, 1990.
[15] K. Souissi, F. Odeh, H.H.K. Tang, and A. Gnudi, *COMPEL*, **13**, 439, 1994.
[16] Y. Apanovich, E. Lyumkis, B. Polsky, A. Shur, and P. Blakey, *IEEE Trans. Computer-Aided Design*, **13**, 702, 1994.
[17] K. Kometer, G. Zandler, and P. Vogl, *Phys. Rev. B*, **46**, 1382, 1992.
[18] G. Zandler, A. Di Carlo, K. Kometer, P. Lugli, P. Vogl, and E. Gornik, *IEEE Electron Device Lett.*, **14**, 77, 1993.
[19] K. Fukuda and K. Nishi, In *Proceedings, Sixth Int. Conf. Simulation of Semiconductor Devices and Processes (SISDEP'95)*, Erlangen, Germany, September 1995.
[20] A. Das and M.S. Lundstrom, *Solid-State Electronics*, **33**, 1299, 1990.
[21] M.A. Alam, M.A. Stettler, and M.S. Lundstrom, *Solid-State Electronics*, **36**, 263, 1993.
[22] C. Huster and M.S. Lundstrom, In *Proceedings, Third Int. Workshop on Computational Electronics*, Portland, OR, May 1994.
[23] N. Goldsmith, L. Henrickson, and J. Frey, *Solid-State Electronics*, **34**, 398, 1991.
[24] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh, *Solid-State Electronics*, **36**, 575, 1993.
[25] Y.-J. Wu. *Multi-band and Two-dimensional Submicron Semiconductor Device Modeling by Direct Solution to the Boltzmann Transport Equation.* PhD thesis, University of Maryland, September 1994.
[26] M.C. Vecchi, D. Venturi, A. Gnudi, and G. Baccarani, In *Proceedings, Int. Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits (NUPAD V)*, Honolulu, HI, June 1994.
[27] R.W. Hockney and J.W. Eastwood. *Computer Simulation Using Particles.* Adam Hilger, New York, 1988.
[28] C. Jacoboni and L. Reggiani, *Rev. Mod. Phys.*, **55**, 645, 1983.
[29] K. Throngnumchai, K. Asada, and T. Sugano, *IEEE Trans. Electron Devices*, **33**, 1005, 1986.
[30] M.V. Fischetti and S.E. Laux, *Phys. Rev. B*, **38**, 9721, 1988.
[31] C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation.* Springer-Verlag, Vienna, 1989.
[32] F. Venturi, R.K. Smith, E.C. Sangiorgi, M.R. Pinto, and B. Riccó, *IEEE Trans. Computer-Aided Design.*, **8**, 360, 1989.
[33] S.E. Laux, M.V. Fischetti, and D.J. Frank, *IBM J. Res. Develop.*, **34**, 466, 1990.
[34] M.V. Fischetti, S.E. Laux, and E. Crabbé, *J. Appl. Phys.*, July 15 1995. (to appear).
[35] R. Stratton, *Phys. Rev.*, **126**, 2002, 1962.
[36] K. Bløtekjær, *IEEE Trans. Electron Devices*, **17**, 38, 1970.
[37] P.A. Sandborn, A. Rao, and P.A. Blakey, *IEEE Trans. Electron Devices*, **36**, 1244, 1989.
[38] M.A. Stettler, M.A. Alam, and M.S. Lundstrom, *IEEE Trans. Electron Devices*, **40**, 733, 1993.
[39] D.L. Woolard, H. Tian, M.A. Littlejohn, K.W. Kim, R.J. Trew, M.K. Ieong, and T.W. Tang, *J. Appl. Phys.*, **74**, 6197, 1993.
[40] A.M. Anile and O. Muscato, *J. Transp. Theory and Stat. Physics*, 1995. (to appear).
[41] E.M. Conwell, In F. Seitz, D. Turnbull, and H. Ehrenreich, editors, *Solid State Physics*, Supplement 9, New York, 1967. Academic Press.
[42] ISE Integrated Systems Engineering AG, Gloriastrasse 35, CH-8092 Zurich. $\mathrm{DESSIS}_{\mathrm{ISE}}$ *Reference Manual Version 1.3*, 1994.
[43] F.J. Blatt. *Physics of Electronic Conduction in Solids.* McGraw-Hill, New York, 1968.
[44] B. Meinerzhagen and W.L. Engl, *IEEE Trans. Electron Devices*, **35**, 689, 1988.
[45] G. Arfken. *Mathematical Methods for Physicists.* Academic Press, New York, second edition, 1970.
[46] A. Lacaita, *Semicond. Sci. Technol.*, **7**, B590, 1992.
[47] C.C.C. Leung and P.A. Childs, *Appl. Phys. Lett.*, **66**, 162, 1995.
[48] W. Hänsch and M. Miura-Mattausch, *J. Appl. Phys.*, **60**, 650, 1986.
[49] A.M. Anile, March 1995. (private communication).
[50] D. Chen, E. Sangiorgi, M.R. Pinto, E.C. Kan, U. Ravaioli, and R.W. Dutton, In *Proceedings, Int. Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits (NUPAD IV)*, Seattle, WA, June 1992.
[51] C. Herring and E. Vogt, *Phys. Rev.*, **101**, 944, 1956.
[52] M.V. Fischetti and S.E. Laux, *Int. Electron Devices Meeting*, December 1995. (submitted).
[53] K. Sonoda, M. Yamaji, K. Taniguchi, and C. Hamaguchi, In *Proceedings, Sixth Int. Conf. Simulation of Semiconductor Devices and Processes (SISDEP'95)*, Erlangen, Germany, September 1995.
[54] C.-S. Yao, J.-G. Ahn, Y.-J. Park, H.-S. Min, and R.W. Dutton, *IEEE Electron Device Lett.*, **16**, 26, 1995.
[55] H. Kosina and S. Selberherr, *IEEE Trans. Computer-Aided Design*, **13**, 201, 1994.
[56] U. Krumbein, D. Yoder, A. Benvenuti, A. Schenk, and W. Fichtner, In *Proceedings, Sixth Int. Conf. Simulation of Semiconductor Devices and Processes (SISDEP'95)*, Erlangen, Germany, September 1995.
[57] R. Thoma, A. Emunds, B. Meinerzhagen, H.-J. Peifer, and W.L. Engl, *IEEE Trans. Electron Devices*, **38**, 1243, 1991.
[58] D. Chen, E.C. Kan, U. Ravaioli, C.W. Shu, and R.W. Dutton, *IEEE Electron Device Lett.*, **13**, 26, 1992.
[59] P. Degond, April 1995. (private communication).
[60] M. Nekovee, B.J. Geurts, H.M.J. Boots, and M.F.H. Schuurmans, *Phys. Rev. B*, **45**, 6643, 1992.