

Numerical modeling of advanced semiconductor devices

by W. Lee
S. E. Laux
M. V. Fischetti
G. Bacarani
A. Gnudi
J. M. C. Stork
J. A. Mandelman
E. F. Crabbé
M. R. Wordeman
F. Odeh

Numerical modeling of the electrical behavior of semiconductor devices is playing an increasingly important role in their development. Examples that pertain to advanced MOSFETs and bipolar transistors are presented to illustrate the importance of taking into account three-dimensional as well as nonequilibrium and nonlocal physical phenomena to effectively characterize the electrical behavior of such devices.

1. Introduction

The extraordinary advances in semiconductor technology in the past four decades have resulted in the mass production of semiconductor devices having characteristic lengths below 1 μm . Developments in photolithography and dry etching techniques have made it possible to define submicron horizontal features on semiconductor wafers. Concurrently, low-thermal-budget processes, such as rapid thermal annealing and plasma-enhanced chemical vapor deposition, have made it possible to achieve junction depths of the order of 10 nm. This miniaturization has necessitated an improvement in computer modeling of the electrical characteristics of small devices.

On one hand, it is becoming increasingly difficult to use one-dimensional or even two-dimensional (2D) models to describe the behavior of many small device structures; three-dimensional (3D) models are necessary to address the geometric dependencies of such structures. On the other hand, today's devices are characterized by relatively large electric fields and relatively sharp spatial variations of the electric fields and carrier concentrations. The variations can occur over distances comparable to the distance between scattering events in a semiconductor crystal. More detailed physical models are therefore necessary in order to facilitate the modeling of associated microscopic transport mechanisms. However, increasing either the geometric dimensionality or the physical rigor of a device model comes at the expense of increased computing times. Despite the rapid decrease in computing costs in recent years, a judicious choice of the device model used to study a particular problem at hand continues to be necessary for efficient use of computing resources.

In this paper, a number of examples of simulations of advanced MOSFETs and bipolar devices are presented to illustrate how numerical modeling can be useful for developing an understanding of the operation of such devices and providing guidance in their design. To examine

©Copyright 1992 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

different aspects of device operation, device models of various levels of sophistication are used in the examples presented. It is our hope that, through these examples, readers will comprehend how simulation tools can be used to address some of the most difficult design problems in both MOSFETs and bipolar devices. We do not present an exhaustive survey of device modeling activities, either within or external to IBM; there exist a number of excellent review papers and books that address various aspects of this rapidly evolving field [1-6]. Instead we focus on the application of advanced numerical device simulation tools to provide insights regarding some of the most difficult current problems in device design.

Some of the simulation tools used in the examples presented are very computationally intensive and may not be practical or economically feasible for day-to-day design applications. Their uses are currently limited but are expected to grow as larger and faster computers become available. By examples of advanced simulation tool usage, we attempt to convey our perspectives on some of the challenges to future device modeling. Many important aspects are not covered. We especially recognize the intimate relationship between process and device models and that between device and circuit models, as covered in the literature [1-7].

In Section 2, we present examples in which use is made of a 3D drift-diffusion simulation program to model trench-bounded MOSFETs and small-emitter bipolar devices. For both types of devices, 3D capabilities are shown to be essential in order to accurately simulate electrical characteristics. In Section 3, we present examples of the application of hydrodynamic models to simulate the impact ionization in MOSFETs as well as bipolar devices. Section 4 covers the use of an advanced Monte Carlo device simulator to obtain detailed microscopic understanding of carrier transport in scaled MOSFETs and bipolar devices. Conclusions and directions for future work are presented in Section 5.

2. Applications of the three-dimensional drift-diffusion model

Currently, most device modeling is based on the drift-diffusion (DD) model, which treats carrier transport as the combined effects of a drift term due to the electric field and a diffusion term due to the spatial distribution of carriers, namely,

$$\mathbf{J}_n = -q\mu_n n \nabla \phi + qD_n \nabla n \quad (1)$$

and

$$\mathbf{J}_p = q\mu_p p \nabla \phi - qD_p \nabla p, \quad (2)$$

where n , p , $\mathbf{J}_n$, $\mathbf{J}_p$, $\mu_{n,p}$, and $D_{n,p}$ are the concentrations, current densities, mobilities, and diffusivities for electrons

and holes, respectively, q is the electron charge, and ϕ is the electrostatic potential. Mobility and diffusivity are assumed to be governed by the Einstein relation, $D = kT/\mu$, where T is the lattice temperature (assumed to be constant).

The DD model can be shown to arise from the second-moment approximation to the Boltzmann transport equation (BTE) for carrier transport. In the latter, the mobility is only a function of the magnitude of the local electric field. Its limitations are discussed in the subsequent sections. A general-purpose 3D device simulator using the DD model was developed in 1980 [1]; however, the widespread use of the three-dimensional DD models to simulate realistic device structures has been limited primarily because of the large memory and the long execution time they require. This has improved considerably in the last several years because of the availability of larger and more powerful computers. In what follows, two examples of the use of the model for modeling small device structures are presented to illustrate the necessity for 3D simulators to resolve various geometrical effects encountered in current as well as anticipated devices.

• Trench-bounded MOSFETs

In this example, design trade-offs related to the rounding of the trench corners of trench-bounded MOSFETs are investigated. A bounding trench may be either a deep trench containing a DRAM storage capacitor, as in a MINT-type structure [8, 9], or a more shallow oxide-filled trench used solely for isolation [10].

Modeling shows that the presence of sharp trench corners improves device electrical behavior compared to that of LOCOS-isolated [11] devices. However, the gate insulator electric field is intensified at the corners, which may present a reliability concern. Rounding the trench corners by modifying device processing conditions represents one means of reducing the gate insulator field.

A view of a MOSFET bounded by trenches having rounded corners is depicted in **Figure 1**. The gate conductor wraps around the corners, accentuating their influence. The gate sidewall oxide and a portion of the oxide used to fill the trench have been cut away closest to the observer to reveal the gate oxide and more of the source drain (S/D) diffusion in the underlying silicon substrate. The other S/D diffusion is mostly covered by the gate conductor.

Three-dimensional device modeling using FIELDAY [1, 12], based on the DD model, has been used to quantify the trade-off between peak insulator electric field and device electrical characteristics as a function of corner radius of curvature. Since the devices of interest have deep-submicron features in each dimension, the use of 3D modeling was essential for accurate prediction of electrical

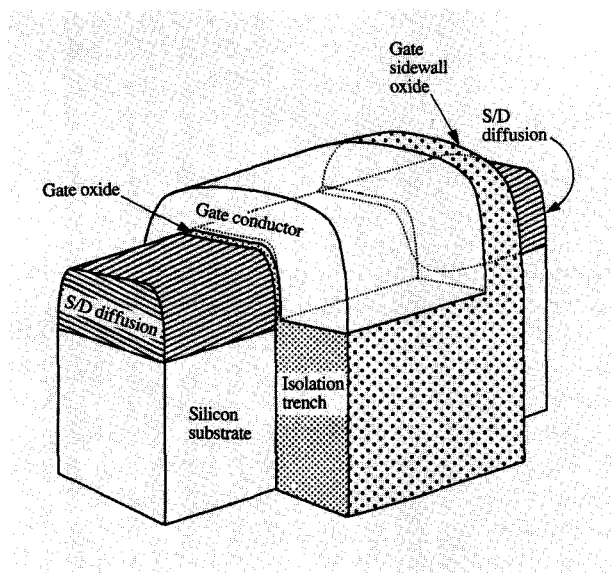


Figure 1

Cutaway view of trench-bounded MOSFET having rounded corners.

behavior. A 2D model would fail to account for interactions with parts of the structure outside of the modeled plane. For example, a 2D model in the vertical plane that cuts through both diffusions, midway across the silicon mesa, would exclude the effects of the trench corners; a vertical plane midway between diffusions would account for trench corner effects, but would contribute no information about the presence of the adjacent diffusions, making the device appear to have an infinite diffusion-to-diffusion spacing (channel length). Finite-element meshes constructed to discretize the 3D structures contain from 25 000 to 40 000 nodes. FIELDAY solves for the carrier concentrations and potentials at each of these nodes, self-consistently with current continuity. Solution of the drift-diffusion equations typically takes 15 to 30 CPU minutes per electrical bias point on an IBM Model 3090™/600E computer running as a uniprocessor, depending on the number of nodes and the degree of nonlinearity encountered.

Figure 2 is a cutaway view of a finite-element mesh used for modeling a trench-bounded MOSFET having a 50-nm corner radius of curvature. Meshes have been generated for modeling corner radii of curvatures ranging from zero to 50 nm, for devices having a 10-nm gate insulator thickness.

The maximum electric field in the gate insulator occurs at the trench corner under the edge of the gate conductor which overlaps the source diffusion, when the gate voltage

is maximum and the source voltage is zero. Modeling predicts that the high electric field in the gate insulator is significantly reduced when the corner radius of curvature is of the order of the gate insulator thickness (**Figure 3**).

For a surface-channel device, the polarity of the electric field in the gate insulator combined with geometric intensification of the electric field at the trench corners results in the corners becoming inverted at a lower gate voltage than the remainder of the silicon surface. The influence of the corners on overall device characteristics increases with decreasing device width; in the deep-submicron regime the influence is significant. Corner rounding alters the inversion electron distribution, affecting device electrical behavior.

Figures 4(a) and 4(b) compare calculated electron minus hole distributions, over the gated silicon surface, for NFETs having sharp and rounded corners, respectively.

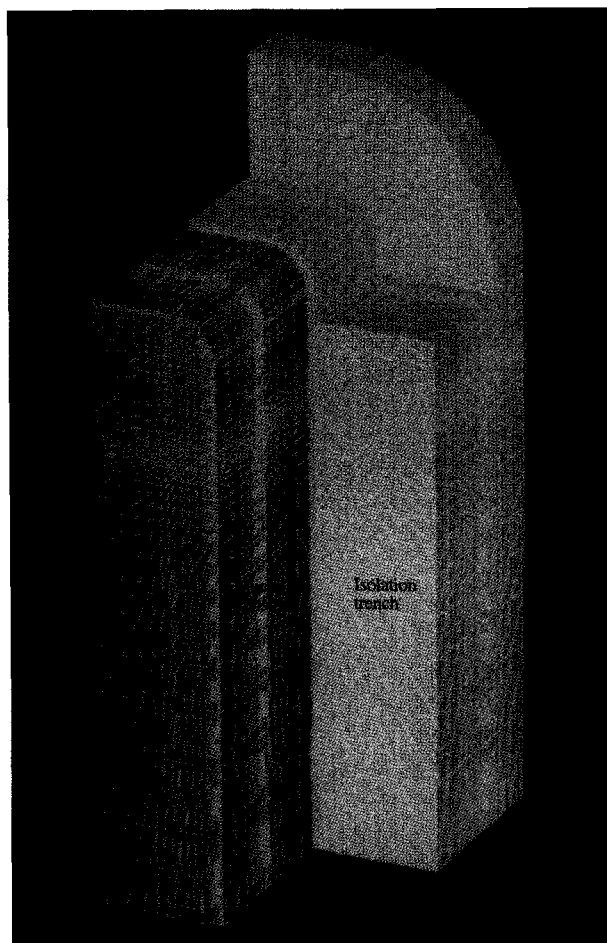


Figure 2

Cutaway view of finite-element mesh for trench-bounded MOSFET model.

Regions in which the electron minus hole concentration is large, such as the source-drain regions, appear bright red. The p-type substrate, where the hole-minus-electron concentration is large, appears bright green. At the zero-gate-voltage condition shown, weak inversion first occurs at the corners, appearing nearly as a filament of electrons at the sharp corners. For the deep-submicron-trench-bounded devices that were modeled, subthreshold current was found to be dominated by corner conduction. Modeling showed that these trench-bounded surface-channel MOSFETs exhibit an enhanced subthreshold slope compared to LOCOS-isolated devices. The improvement in subthreshold slope with decreasing corner radius of curvature is apparent from the modeled transfer characteristics of **Figure 5**.

Another beneficial consequence of the use of sharp trench corners is a reduction in the short-channel effect (the change in threshold voltage arising from a change in channel length). The improvement results from a reduced penetration of the drain field to the source barrier because an increased fraction of drain field lines terminate on the

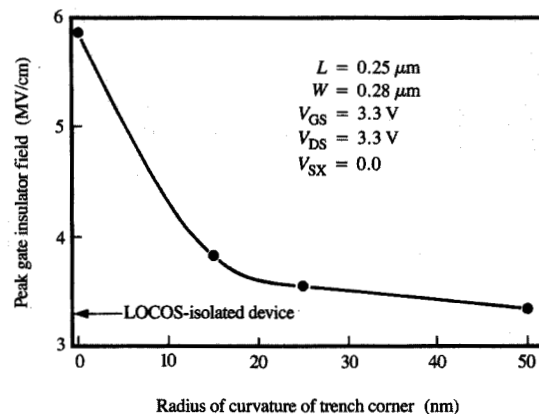
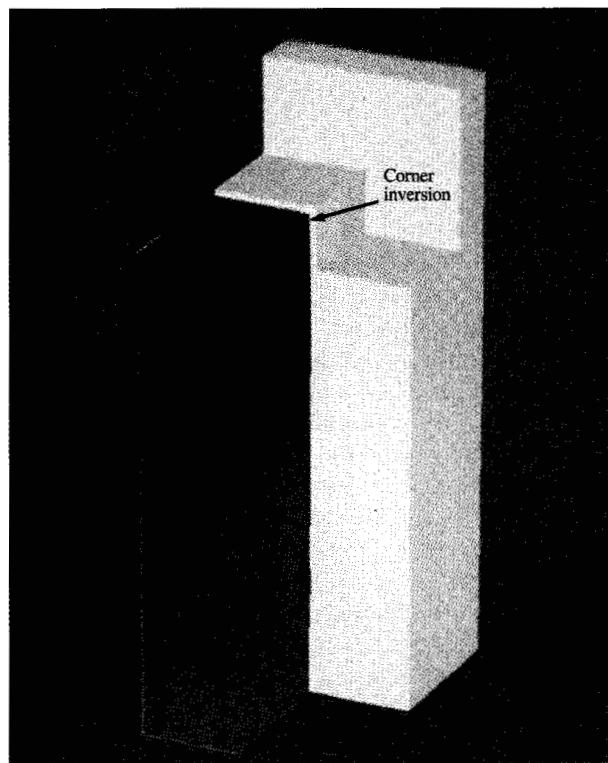
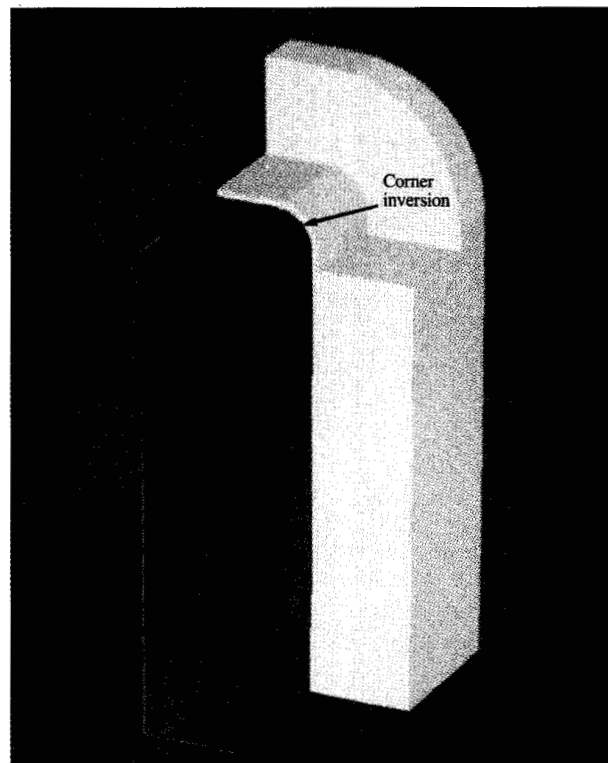


Figure 3

Calculated dependence of gate insulator field on radius of curvature of trench corner.



(a)



(b)

Figure 4

Calculated distributions for electrons minus holes on silicon surface at the onset of corner inversion. Red regions are those in which the electron concentration is greater than the hole concentration. Green regions are those in which the hole concentration is greater than the electron concentration. Black regions are essentially depleted of carriers. For (a), trench corners were assumed to be sharp. For (b), the radius of curvature of the trench corners was assumed to be 50 nm, leading to a more spread-out inversion region.

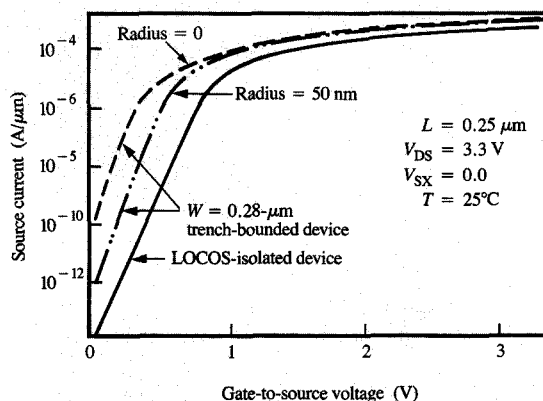


Figure 5

Calculated n-MOSFET transfer characteristics.

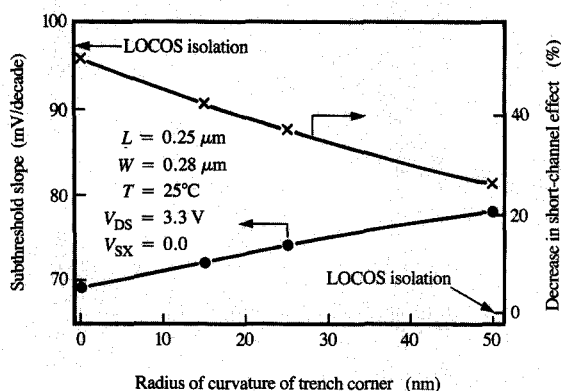


Figure 6

Calculated dependence of subthreshold slope and the short-channel effect on trench corner radius of curvature.

gate conductor at the corners. The calculated improvement in the short-channel effect as a function of corner radius is shown in **Figure 6**, along with a quantification of the modeled subthreshold slope discussed previously.

At those trench corner radii of curvature required to realize a significant reduction in gate insulator field, trench-bounded devices still enjoy substantial advantages over LOCOS-isolated devices. Superior behavior with regard to subthreshold slope and the short-channel effect permit the achievement of a lower threshold voltage for a given off-

current. A lower threshold voltage results in an increased device overdrive ($V_T - V_{GS}$) and an increased on-current; therefore, the MOSFETs should be able to switch more rapidly without sacrificing device size or standby power. Thus, 3D device modeling demonstrates that trench corner rounding should provide a means for reducing the gate insulator electric field to a reliable level while allowing much of the enhancement in deep-submicron-trench-bounded device behavior to be realized.

• Small-emitter bipolar devices

As the downscaling of silicon bipolar transistors progresses toward emitter sizes smaller than $0.5 \mu\text{m}$ [13, 14], the influence of 3D effects on their electrical characteristics is expected to become more pronounced. Various small-emitter effects have been observed experimentally [13]. However, it remains unclear whether some of the observed effects are unique to certain processes or are of a more general nature.

In the next example described, the 3D FIELDAY model [1, 12] has been used to investigate the dc, ac, and punchthrough characteristics of npn bipolar devices having small emitters. The vertical profiles of the devices were maintained constant so that only geometrical effects would affect the resulting differences in their electrical characteristics. It should be noted that there have been only a few reports on 3D simulations of silicon bipolar devices, and none have dealt with the scaling of emitter size [15, 16].

Figure 7(a) shows a schematic cross section of a typical silicon bipolar transistor having an emitter surrounded by a heavily doped extrinsic base region (singly hatched) and a base electrical contact (doubly hatched). To reduce computation time, a quadrant (bounded by the dashed lines) of the physical device structure is usually simulated by invoking symmetry boundary conditions. For comparison purposes, 2D FIELDAY simulations were also performed in the plane of one of the dashed lines. As in the MOSFET modeling in the previous section, finite-element right-angled prismatic 3D meshes containing 15 000 to 40 000 nodes were constructed for 3D simulations of devices with various emitter dimensions. Typically 15 to 30 CPU minutes were required on an IBM RISC System/6000™ workstation to obtain the solution at one bias point. The corresponding 2D meshes contain only 1200 to 4000 nodes. About one CPU minute was required for each bias point. A 3D plot of dopant distribution of a quadrant of a device structure having an emitter area of $0.2 \mu\text{m} \times 0.2 \mu\text{m}$ is shown in **Figure 7(b)**. For comparison purposes, the dimensions of the base contact and the extrinsic base region remain the same for different emitter sizes.

The Gummel characteristics calculated from 2D and 3D simulations of two bipolar transistors having emitter areas

of $0.2\ \mu\text{m} \times 0.2\ \mu\text{m}$ and $1.2\ \mu\text{m} \times 1.2\ \mu\text{m}$ are shown in **Figures 8 and 9**, respectively. The vertical dopant profiles of the two transistors are identical and similar to those reported in [13]. For the larger ($1.2\text{-}\mu\text{m} \times 1.2\text{-}\mu\text{m}$) transistor, the 2D and 3D simulations yield similar results at low biases. At higher biases, the current densities from the 3D simulations are slightly higher because of the lower base resistance in the 3D structure—in which the extrinsic base region surrounds the emitter. For the smaller ($0.2\text{-}\mu\text{m} \times 0.2\text{-}\mu\text{m}$) transistor, current densities calculated from the 3D simulation were higher than those from the 2D simulation. The higher current densities in the 3D structure are primarily due to the contribution from the current flowing into the sidewall area of the base–emitter junction; that area is not negligible in the smaller device; it amounts to half of the bottom junction area. Furthermore, the current gain roll-off at high collector current densities is calculated to be much less severe for the smaller device, as shown in **Figure 10**. This result is primarily attributed to the 3D spreading of electron carriers outside the intrinsic region directly under the emitter, which advances the onset of the Kirk effect [17] to a higher current density level.

The cut-off frequency f_T , obtained from the 2D and 3D simulations of the same pair of devices, is shown in **Figure 11** as a function of operating collector current density. At low current densities, the f_T calculated for the smaller device is lower than that calculated for the larger device, because of the higher ratio of the extrinsic to intrinsic base–collector capacitance for the smaller device. The peak f_T of the smaller device is also lower than that of the larger one. In addition to the higher ratio of extrinsic to intrinsic capacitance in the smaller device, the 3D spreading of the electron charge also leads to a longer average transit time, thus a lower value of f_T . The longer transit time is due to the effectively longer average transit distance for carriers before they are collected by the collector in the small 3D structure. Furthermore, the f_T roll-off of the smaller device is much less severe at high current densities, which is again due to 3D current spreading and is consistent with its dc behavior, as described in the previous paragraph.

Figure 12 shows the calculated 3D equidensity surface for electrons in the $0.2\text{-}\mu\text{m} \times 0.2\text{-}\mu\text{m}$ device. Superimposed are the transparent cutaway planes for the dopant density to indicate the location of the emitter and the intrinsic base region. As can be seen, a significant number of electrons are located outside the intrinsic device region.

Next, 2D and 3D simulations were performed to investigate the geometry dependence of the punchthrough characteristics. The results are shown in **Figure 13**. In these punchthrough simulations, the base and collector were assumed to be at ground potential, while the emitter

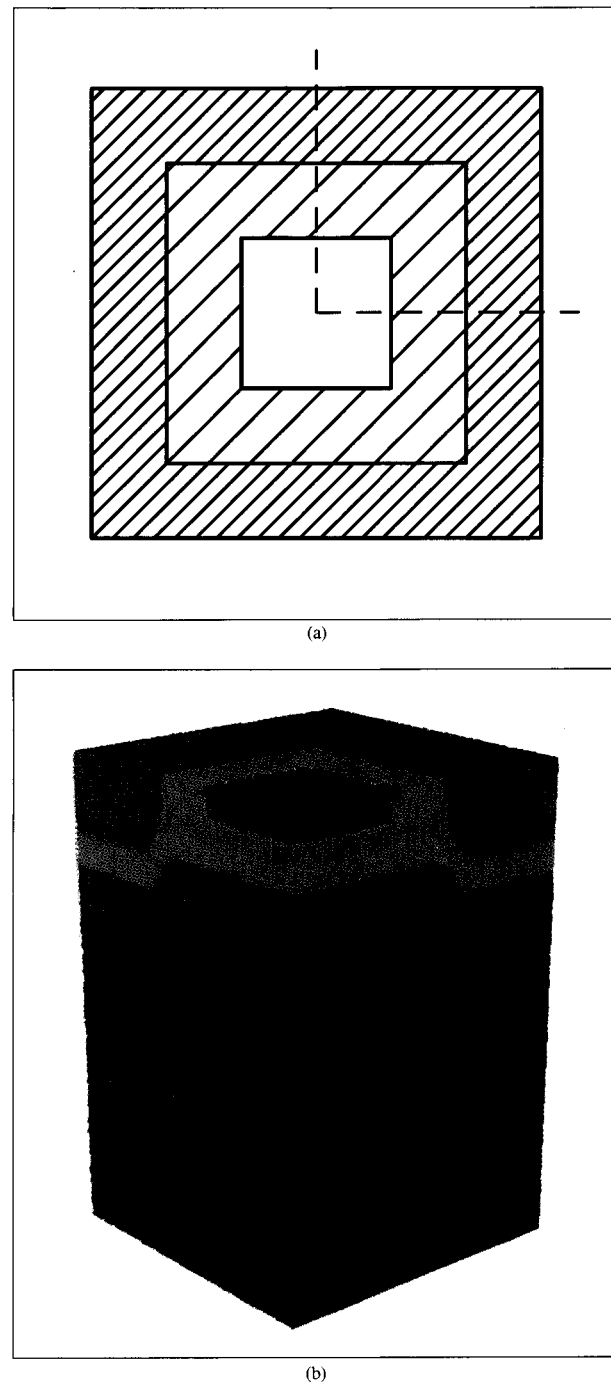


Figure 7

(a) Plan view of a typical silicon bipolar transistor with emitter (center region) surrounded by an extrinsic base region (singly and doubly hatched) and a base electrical contact (doubly hatched). (b) Plot of the dopant distribution of a device structure having an emitter area of $0.2\ \mu\text{m} \times 0.2\ \mu\text{m}$. Only a quadrant is shown. The cyan coloration indicates a p-type dopant density from 2×10^{18} to $1 \times 10^{19}\ \text{cm}^{-3}$, and the purple coloration indicates a p-type dopant density above $1 \times 10^{19}\ \text{cm}^{-3}$; n-type dopant density is indicated by the following colored intervals: red, above $5 \times 10^{17}\ \text{cm}^{-3}$; brick, from 1×10^{17} to $5 \times 10^{17}\ \text{cm}^{-3}$.

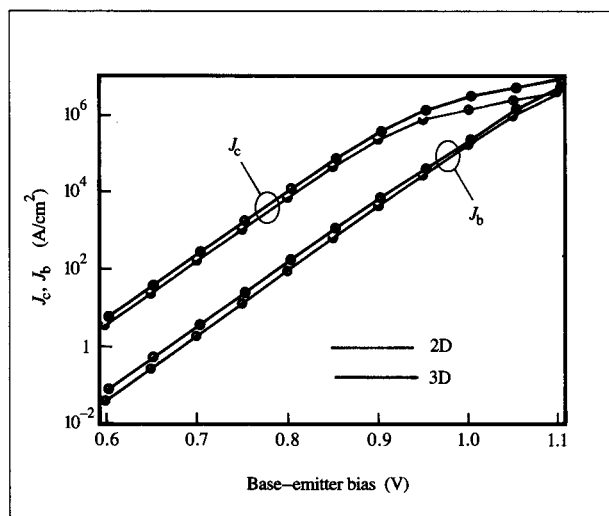


Figure 8

Calculated Gummel characteristics of a silicon bipolar transistor having an emitter area of $0.2 \mu\text{m} \times 0.2 \mu\text{m}$. Results for 3D and 2D simulations are shown by the red and blue lines, respectively. Corresponding collector current densities are designated as J_c and J_b .

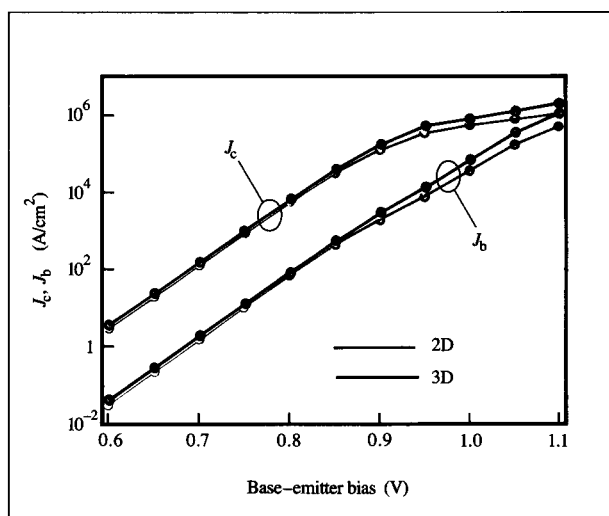


Figure 9

Calculated Gummel characteristics of a silicon bipolar transistor having an emitter area of $1.2 \mu\text{m} \times 1.2 \mu\text{m}$. Results for 3D and 2D simulations are shown by the red and blue lines, respectively.

contact was assumed to be biased with a positive voltage in order to achieve reverse biasing of the base-emitter junction. As is evident from the figure, the calculated punchthrough current density at the same bias is about two

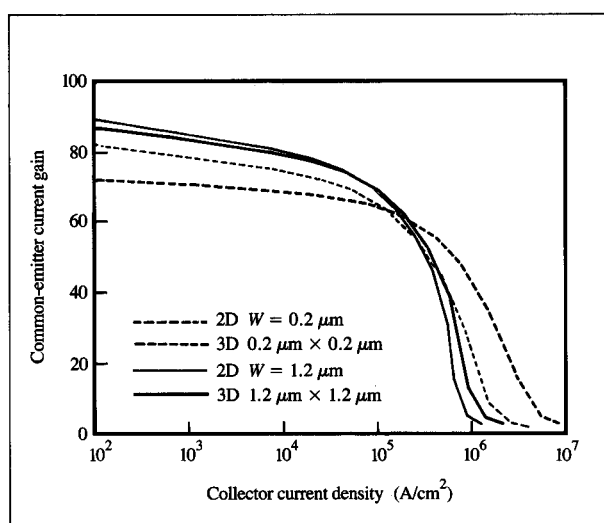


Figure 10

Calculated common-emitter current gain as a function of collector current density. The emitter width w is indicated for the 2D cases.

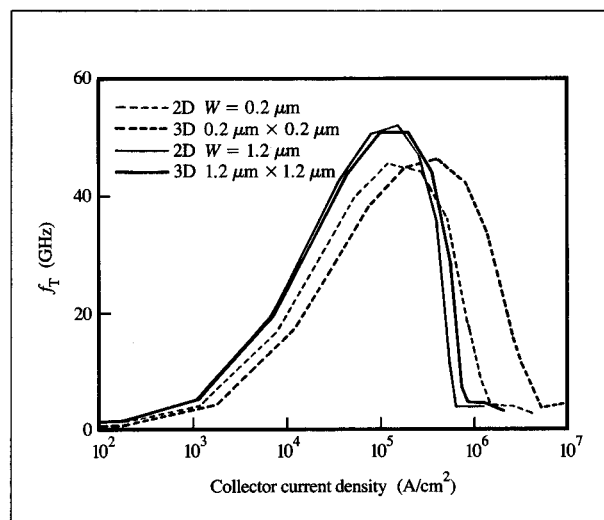


Figure 11

Calculated common-emitter cutoff frequency f_T as a function of collector current density.

orders of magnitude lower for the smaller device. A close examination of the field distribution reveals that there is a significant amount of charge sharing between the extrinsic base-collector space-charge regions. As a result, punchthrough is always expected to begin in the interior part of the intrinsic base (Figure 14), which is the

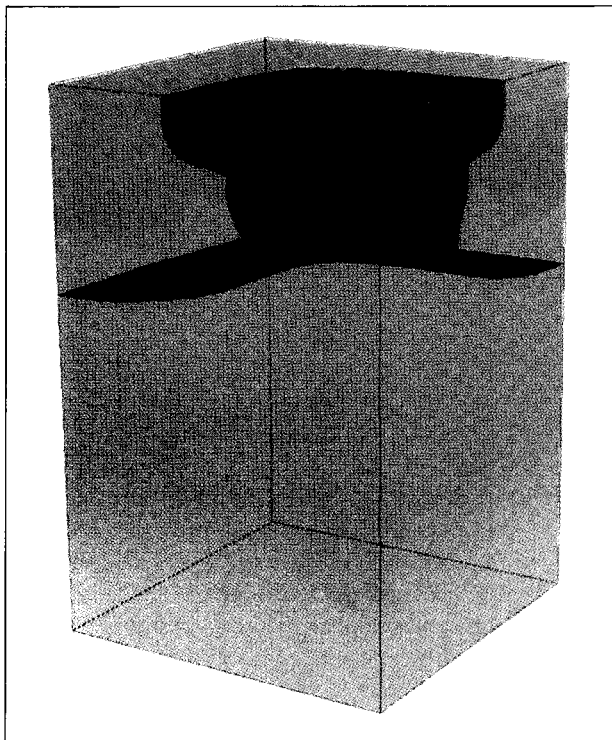


Figure 12

Calculated 3D electron equidensity surface (in blue) and cutaway planes of the doping profile of the bipolar device of Figure 7, at a base-emitter bias of 0.9 V. The base region is shown in red and purple. The bright green and light green areas indicate n-type doping above and below $5 \times 10^{17} \text{ cm}^{-3}$, respectively. The calculations indicate that significant numbers of electrons are injected into the extrinsic base region. The focusing effect on electrons near the base-collector junction is attributed to the lateral field of the extrinsic base-collector junction.

reason why smaller devices tend to have lower punchthrough current densities at the same bias.

From the above comparisons between 2D and 3D simulations of dc, ac, and punchthrough characteristics, it is apparent that 3D simulations are essential for predicting the electrical performance of bipolar devices having emitter sizes smaller than $0.5 \mu\text{m}$. The use of 2D simulations yields results quantitatively similar to those of 3D simulations for devices with emitter widths $\geq 1 \mu\text{m}$. However, as emitter widths decrease to below $0.5 \mu\text{m}$, significant quantitative differences exist between the 2D and 3D simulations.

3. Applications of hydrodynamic models

• Two-dimensional modeling of the substrate current in MOSFETs

As discussed in the Introduction, current semiconductor devices are characterized by relatively large electric fields

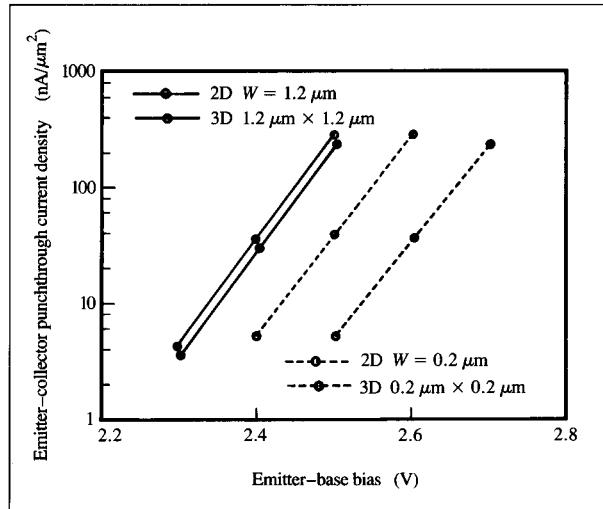


Figure 13

Calculated emitter-collector punchthrough current density, as a function of emitter-base bias.

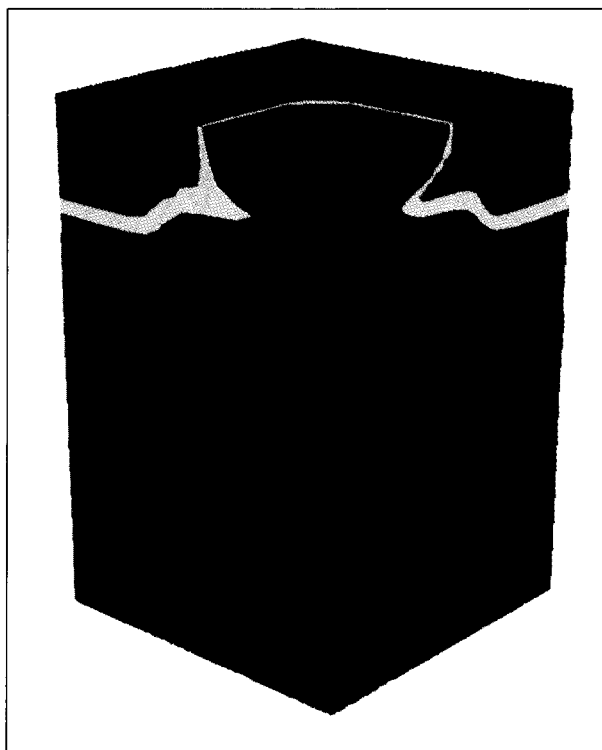


Figure 14

Calculated 3D contour plot of hole density distribution at an emitter-base bias of 2.5 V. Blue: below 10^{10} cm^{-3} . Green: from 10^{10} to 10^{15} cm^{-3} . Yellow: from 10^{15} to 10^{18} cm^{-3} . Orange: from 10^{18} to 10^{19} cm^{-3} . Red: above 10^{19} cm^{-3} .

as well as relatively sharp spatial variations of the electric fields and carrier concentrations. The carriers in these devices can be "heated" by applied electric fields, resulting in a carrier energy distribution that deviates significantly from its equilibrium shape. This violates a basic assumption in the DD model. Qualitatively, the hydrodynamic approach extends beyond the DD approach by allowing carrier mobilities to be functions of the average carrier energy rather than the local electric field, and allowing the carrier temperature to be different from that of the lattice. The fundamental framework of the hydrodynamic approach is outlined below for completeness, although this topic has been treated elsewhere in the literature [5, 6, 18].

In the hydrodynamic model, the fundamental equations for electrons in steady state are as follows:

$$U - \frac{1}{q} \operatorname{div}(\mathbf{J}_n) = G_{II}, \quad (3)$$

$$\mathbf{J}_n - \frac{\tau_{pn}}{q} (\mathbf{J}_n \cdot \nabla) \frac{\mathbf{J}_n}{n} = q\mu_n \left[\frac{k_B T_n}{q} \nabla(n) + n \nabla \left(\frac{k_B T_n}{q} - \phi \right) \right], \quad (4)$$

and

$$\begin{aligned} -\operatorname{div} \left[\kappa_n \nabla(T_n) + \frac{\mathbf{J}_n}{q} (w + k_B T_n) \right] \\ = \mathbf{E} \cdot \mathbf{J}_n - n \frac{w - w_0}{\tau_{wn}} - Uw - E_{th} G_{II}, \end{aligned} \quad (5)$$

where n and $\mathbf{J}_n$ denote the electron and the current densities, respectively, and w denotes the electron average energy; the latter is related to the mean velocity v and the effective electron temperature T_n by the expression

$$w = \frac{1}{2} m v^2 + \frac{3}{2} k_B T_n. \quad (6)$$

Moreover, w_0 represents the equilibrium energy, and ϕ and $\mathbf{E}$ are the electrostatic potential and field, respectively; the latter are governed by the Poisson equation. Expression (6) gives the average energy as the sum of the contributions from the average drift velocity and the random motion around that average. The terms G_{II} and E_{th} are discussed in a later paragraph.

These equations are derived by the "moments" method applied to the Boltzmann transport equation under the following simplifying assumptions: 1) the temperature tensor (defined as the second moment of the particle distribution function around its mean) reduces to a scalar; 2) the hierarchy of moments is closed by relating the *heat flow* of the electron gas $\mathbf{Q}$ to the electron temperature via $\mathbf{Q} = -\kappa_n \nabla(T_n)$, where κ_n is the *thermal conductivity*, which is given by the Wiedmann-Franz law $\kappa_n = (5/2 + c)(k_B/q)^2 \sigma_n T_n$; and 3) the effects of collisions

on momentum and energy are modeled by the relaxation time approximation. In the fundamental equations, τ_{pn} and τ_{wn} denote the momentum and energy relaxation times, and $\mu_n = (q/m)\tau_{pn}$. Usually the mobility μ_n is modeled as a decreasing function of the average energy, because of an observed increase in scattering rates with particle energy.

We now discuss some of the features that distinguish the hydrodynamic from the DD approach. The first is the introduction of the energy balance equation (5). On the right-hand side we recognize 1) the forcing term $\mathbf{E} \cdot \mathbf{J}_n$, which represents the energy absorbed by the electrons from the electric field per unit time and volume; 2) the energy $n(w - w_0)\tau_{wn}$ delivered by the electrons to the lattice per unit time and volume because of collisions with optical phonons; and 3) the energy lost in the recombination processes Uw . The algebraic sum of the three terms is equal to the spatial variation of the energy flow at the left-hand side, which is the sum of the heat flow and the transported energy. Because of the finite nature of the energy-relaxation time, the mean energy w and the electron temperature T_n turn out to be larger than their equilibrium values in regions of large electric fields. The second feature stems from Equation (4). Apart from the nonlinear convective term $(\tau_{pn}/q)(\mathbf{J}_n \cdot \nabla)(\mathbf{J}_n/n)$, there are two main differences from the DD model. One is the existence of a *thermoelectric field*, the gradient of $k_B T_n/q$, which provides an additional driving force that accounts for the tendency of electrons to diffuse from "hot" toward "cold" regions. The other is the relation between the diffusion coefficient and the mobility, given by

$$D_n = \frac{k_B T_n}{q} \mu_n, \quad (7)$$

which involves the electron temperature T_n rather than the lattice temperature. Thus, in hot regions, where T_n is large, the hydrodynamic model predicts a larger diffusion than does the DD model, reflecting the increase in the random velocity over its equilibrium level.

Another phenomenon that stems from the energy equation and from the assumptions mentioned above regarding the mobility is *velocity overshoot*. This phenomenon can be seen most easily in a one-dimensional setting in which the electric field increases in the direction of motion. A simple consequence of the energy balance equation is that under homogenous conditions the average energy w is less than the value corresponding to the local field. Thus, since the mobility is a decreasing function of w , it follows that the hydrodynamic velocity is larger than the velocity based on a mobility that is a function of the local electric field, as is commonly assumed in the DD model (for details see [19]).

A final remark on the hydrodynamic approach: The numerical aspects of the hydrodynamic approach are somewhat more delicate than those of the DD approach

because of the following: 1) the increased *stiffness* of the momentum equation compared to Equations (1) and (2); 2) the complications introduced by the convective term, permitting the achievement of extreme gradients in velocity and electron density; and 3) the roughness of the source term $\mathbf{E} \cdot \mathbf{J}$ in the energy equation, because of the difficulty of obtaining a smooth representation of the current density vector $\mathbf{J}$. These issues have been addressed in some specialized papers (e.g., [20]), in which it is shown that they can be taken into account by means of a generalized Scharfetter–Gummel scheme coupled to careful iteration methods.

The modeling of substrate currents in MOSFETs due to impact ionization is of importance in device design. Several papers have recently appeared on this subject [21, 22], offering a variety of methods that range from the use of semi-analytical models, to the direct application of the energy balance equation, to, finally, the combination of the DD or the energy-balance equations with Monte Carlo simulations. These methods aim at extracting more detailed information regarding the high-energy tail of the distribution.

If the DD model is assumed to be applicable, impact ionization is usually treated in the spirit of the so-called *lucky-electron* model, which derives its main concept from the original work of Shockley [23]; there are obvious limitations in using this approach for short devices because of its neglect of effects caused by fast variations of the electric field. Several proposals have been made for incorporating impact ionization in the hydrodynamic equations in a more physically sound way; for example, in [21] the ionization coefficient is represented by an exponential in the homogeneous electric field associated with the electron temperature, while [22] uses a generalization of the lucky-electron concept coupled to the hydrodynamic model.

Another approach, proposed in [24], has been adopted in the simulations presented later in this section. A short description follows: One assumes that the even part of the electron energy distribution can be described by a *heated Maxwellian* function of the form $f(E) \propto \exp(-E/k_B T_n)$. The microscopic ionization scattering rate is then calculated assuming a Coulomb interaction potential between two colliding electrons, one in the conduction band and the other in the valence band. The resulting generation rate turns out to be a function of the electron temperature, having the form

$$G_{II} = nY(u) = \frac{1}{2\tau_0} [\sqrt{u/\pi} \exp(-1/u) - \operatorname{erfc}(1/\sqrt{u})], \quad (8)$$

where τ_0 is a normalization time constant, and $u = (k_B T_n)/E_{th}$ is proportional to the electron mean energy normalized by the ionization threshold energy E_{th} . The

effects due to sharply varying electric fields are incorporated by means of the T_n dependence, via the solution of the energy balance equation. The generation rate enters the continuity equation [Equation (3)] and the energy balance equation [Equation (5)]. The term $E_{th} G_{II}$ in (5) represents the loss of energy per unit volume and time because of the ionization processes. The coefficients τ_0 and τ_{wn} can be considered as adjustable parameters used to fit experimental results, as shown below.

In the remaining paragraphs in this section, we present the results of the simulation of the behavior of short n-channel MOSFETs, taking into account impact ionization, and we compare the calculated substrate currents with measured data. The simulated structures were the n-channel silicon devices of the 0.25- μm CMOS technology presented in [25], characterized by an oxide thickness $t_{ox} = 7$ nm, a junction depth $x_j = 0.1$ μm , and a substrate doping level of 10^{17} cm^{-3} . The MOSFETs were assumed to have a different effective channel length L , ranging from 0.35 μm to 4 μm . The simulations were carried out with the hydrodynamic version of the two-dimensional device analysis program HFIELDs [26]. The latter solves the associated electron hydrodynamic equations, including the generation term G_{II} in the form of Equation (5), and does so self-consistently with the Poisson equation. In an n-channel device, for conditions far from the drain-bulk breakdown, holes are not expected to contribute appreciably to the total ionization rate, and hence are treated in the drift-diffusion approximation. We have used the mobility model [27], defined by

$$\mu_n(w) = \frac{\mu_0}{1 + \alpha(w - w_0)}, \quad (9)$$

in which

$$\alpha = \frac{\mu_0}{q\tau_{wn}v_{sat}^2}, \quad (10)$$

with $\mu_0 = \mu_{00}/(1 + E_{\perp}/E_{ref})$, where $E_{\perp}$ is the electric field perpendicular to the Si–SiO₂ interface, accounting for surface mobility degradation. The turn-on characteristics for different channel lengths with $V_{DS} = 50$ mV (Figure 15) were used to obtain values of the parameters μ_{00} and E_{ref} . The source/drain contact resistances, whose effect is non-negligible in the shortest simulated device ($L = 0.35$ μm), were accounted for by means of external lumped resistances in series with the source/drain contacts. To speed up the fitting procedure for obtaining the contact resistance, use was made of a semi-analytical approach [28].

The calculated high-drain voltage turn-on characteristics (at $V_{DS} = 2.5$ V) showed good agreement with experimental data [29] when $v_{sat} = 9.2 \times 10^6$ cm/s was used in Equation (10), as indicated in Figure 16. We have

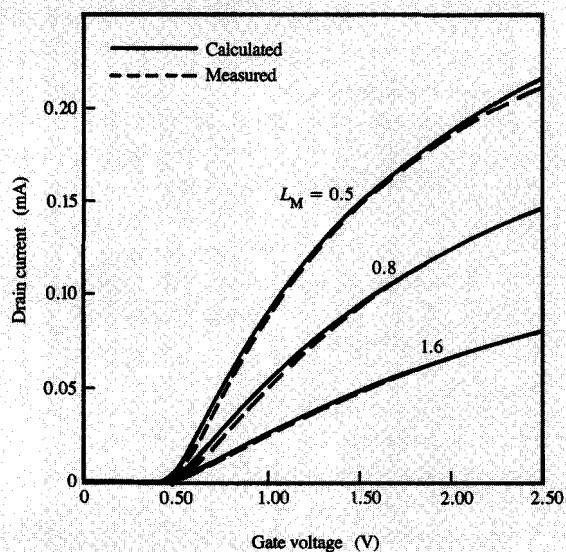


Figure 15

Calculated and measured turn-on characteristics (at $V_{DS} = 50$ mV) for the silicon n-MOSFETs discussed in the text, for three different channel lengths (L_M is the associated gate mask length in μm).

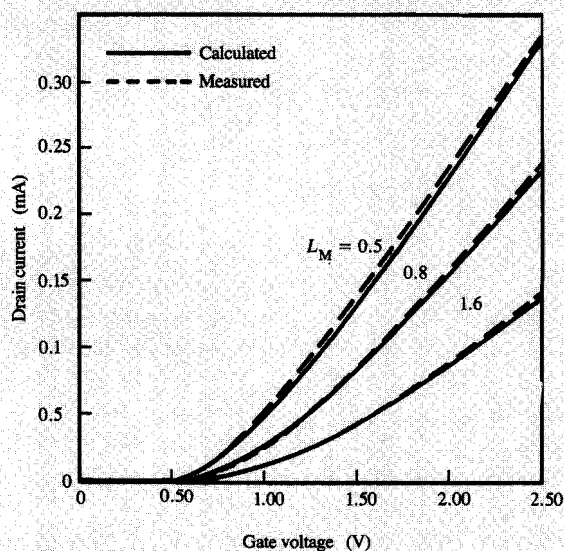


Figure 16

Calculated and measured high-drain voltage turn-on characteristics (at $V_{DS} = 2.5$ V) for the n-MOSFETs discussed in the text, for three different channel lengths.

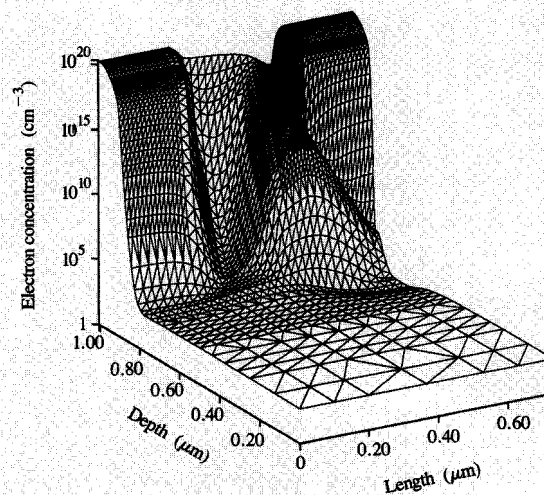


Figure 17

Calculated electron concentration in simulated $0.35\text{-}\mu\text{m}$ -channel-length n-MOSFET (at $V_{GS} = 1.5$ V and $V_{DS} = 3.0$ V).

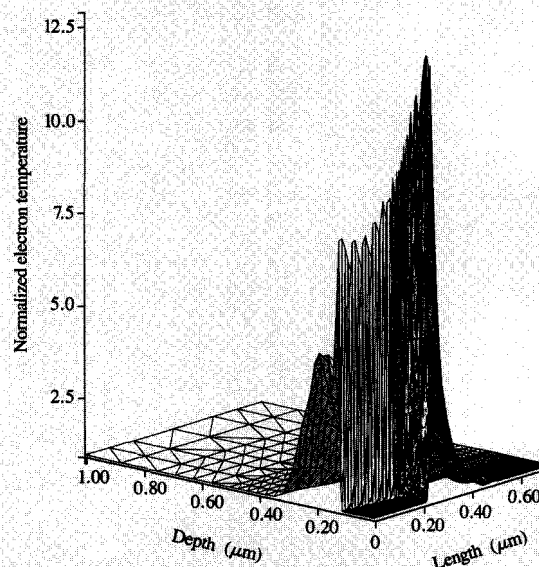


Figure 18

Normalized electron temperature T_e/T_L ($T_L = 300$ K) in simulated $0.35\text{-}\mu\text{m}$ -channel-length n-MOSFET (at $V_{GS} = 1.5$ V and $V_{DS} = 3.0$ V).

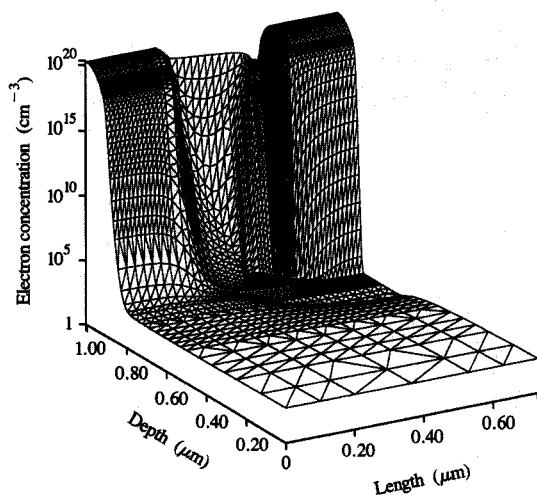


Figure 19

Corresponding electron concentration obtained with the DD model.

observed that the effect of velocity overshoot on the drain current is quite small. In fact, the overshoot generally depends on the relaxation time τ_{wn} ; however, even in the shortest device, changing τ_{wn} within the typical values in the literature does not significantly affect the drain current. This can be accounted for by considering that the velocity overshoot is due primarily to the spatial variation of the electric field. In MOSFETs, the largest variation of the field occurs near the drain, while the drain current is controlled by the portion of the channel at the source side.

Typical calculated characteristics of the electron concentration and normalized electron temperature (T_n/T_L) for the 0.35- μm -channel-length MOSFET ($V_{GS} = 1.5\text{ V}$, $V_{DS} = 3.0\text{ V}$) are shown in **Figures 17 and 18**, respectively. For comparison, the corresponding electron concentration calculated using the DD model is shown in **Figure 19**. The temperature clearly exhibits a ridge around the high-field region near the drain, and suddenly drops as the neutral drain region is approached. The ripple on the bulk side of the drain junction is attributed to the aforementioned roughness of the $\mathbf{E} \cdot \mathbf{J}$ term, and to the difficulty of accurately resolving the very rapid, associated variation in temperature. Comparing **Figures 18 and 19** clearly shows the effect of the increased electron diffusivity that results from the use of the generalized Einstein relation [Equation (7)] in the hydrodynamic model; its use produces an electron concentration around the drain junction that is much higher than that obtained with the DD approach.

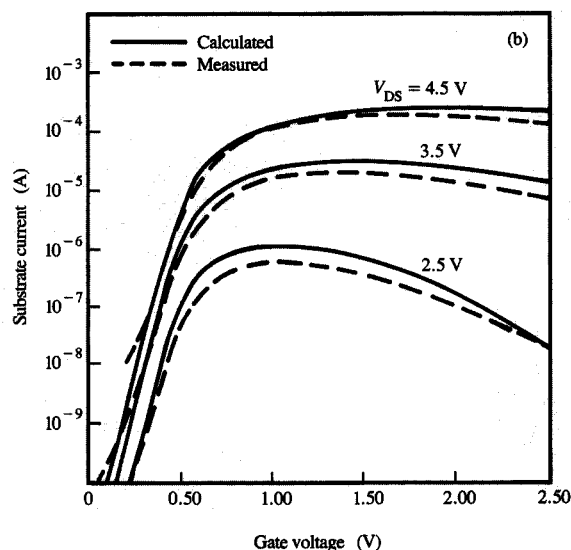
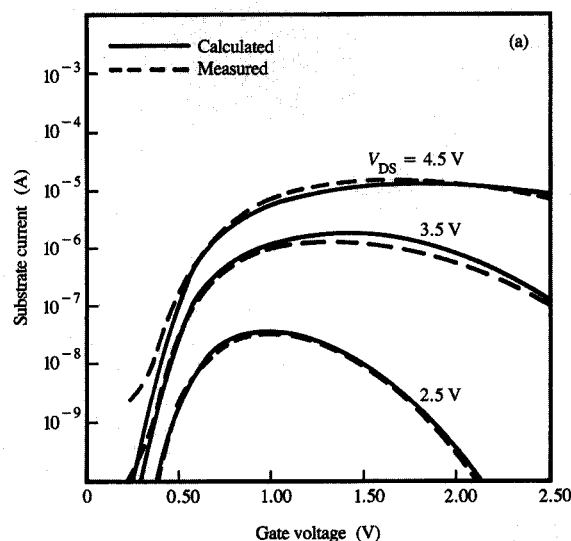


Figure 20

Calculated (using hydrodynamic model) and measured substrate currents for the n-MOSFETs discussed in the text, for two channel lengths: (a) 4.0 μm ; (b) 0.5 μm .

Calculated and measured substrate currents for different bias conditions and channel lengths are shown in **Figure 20**. The two adjustable parameters τ_0 and τ_{wn} were obtained by comparing the measured and the calculated substrate currents for the longest available channel length and the lowest source-to-drain voltage ($L = 4\text{ }\mu\text{m}$, $V_{DS} = 2.5\text{ V}$). The values thus obtained were $1/\tau_0 = 7 \times 10^{15}\text{ s}^{-1}$

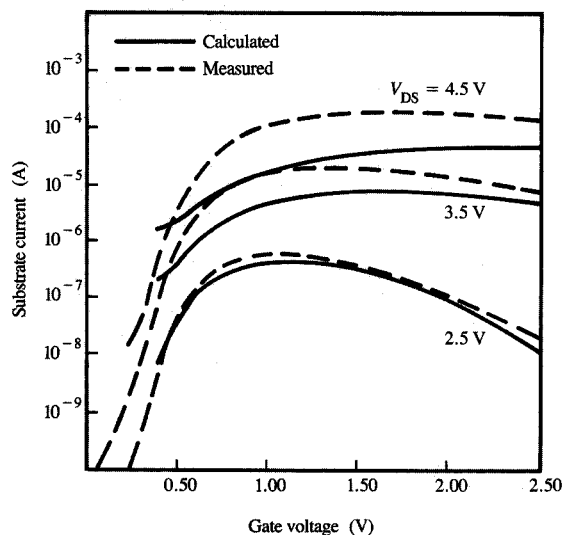


Figure 21

Calculated (using DD model) and measured substrate currents for the 0.5-μm-gate-mask-length case.

and $\tau_{wn} = 0.1$ ps. A suitable value for the ionization threshold energy was found to be $E_{th} = 2.0$ eV, which is quite close to the 1.8 eV extracted experimentally [30]. Figure 20 shows that substrate currents can be predicted with sufficient accuracy for all of the channel lengths considered down to $0.35 \mu\text{m}$, for drain voltages up to at least 4.5 V. For comparison we used a DD version of HFIELDS with an electric-field-dependent impact ionization model that was fitted at a small value of V_{DS} in order to compute the substrate current. **Figure 21** shows the main results obtained—in poor agreement with experimental data, especially for high biases.

In summary, in this section we have briefly described some of the physically relevant aspects of the hydrodynamic model for simulating the behavior of semiconductor devices. We have remarked on some impact ionization models which are physically consistent with the hydrodynamic formulation of carrier transport. A combined hydrodynamic and impact ionization model [resulting in Equation (8)], with two adjustable parameters, was used to simulate the bulk currents of short n-channel MOSFETs. Good agreement was found for a large range of channel lengths and bias conditions.

• One-dimensional modeling of impact ionization in bipolar devices

The performance of a current bipolar transistor is closely related to its carrier transit time. In the past, the base

transit time was the dominant contribution to the total delay, because of the use of relatively large base widths and relatively slow, diffusion-dominated, charge transport. Recent developments in silicon bipolar technology, however, have made it possible to produce bipolar transistors having base widths smaller than 50 nm. Recently, the addition of a suitably profiled Ge distribution within the Si base has made it possible to develop a built-in field which may be as large as 10^4 V/cm, thereby enhancing the average velocity of the carriers [31]. As a result, the base transit time can be kept below 1 ps, no longer constituting the major contribution to the total delay of the carriers. Thus, the collector transit time is emerging as the limiting factor of future-generation devices [32].

One of the major problems associated with the scaling of bipolar transistors is the impact ionization effect which occurs within the collector space-charge region [33]. As the width of the latter is reduced in accordance with scaling rules, the electric field is expected to increase by the scaling factor, leading to increased impact ionization. Holes generated within the collector space-charge region drift to the base neutral region and thereby reduce the base current. For large current-gain devices, the base current eventually reverses, and instability occurs in common-emitter configurations.

In this section, we address the problem of modeling the impact ionization within the collector space-charge region of a bipolar transistor. An analysis is carried out assuming a simplified version of the energy balance equation of the previous section, which, however, retains the ability to account for nonlocal effects [34]. While a complete 2D device simulation based on the hydrodynamic approach is quite possible, the simplified analysis has the distinct advantage of providing a clear indication of the key factors affecting impact ionization.

The one-dimensional form of the energy balance equation can be expressed as

$$\frac{d}{dx} \left(\frac{5}{3} n v w \right) = -q n v E - n \frac{w - w_0}{\tau_w}, \quad (11)$$

where n , v , w , w_0 , and τ_{wn} are as defined in the previous section, and E is the electric field. Two simplifying assumptions are implied by (11): 1) The heat flow is neglected; and 2) the convective energy is assumed to be negligible compared with the thermal energy. These assumptions account for the difference between expressions (11) and (5). The second assumption is fully justifiable under static conditions [34], under which convective energy never exceeds 6% of the total energy. On the other hand, when sudden changes of the field occur over a short distance, such as at the base-collector edge of the space-charge region, the convective energy may become quantitatively more important. This modifies the

exact value near the boundary, but has a negligible effect on the position and magnitude of the maximum electron energy, the key parameter affecting impact ionization. Dropping the convective term implies that temperature and energy become equivalent and interchangeable, simplifying the analysis considerably.

In what follows, it was further assumed that the product $v\tau_{wn}$ is nearly constant. Such an assumption was justified by the fact that carriers travel across most of the collector space-charge region at saturated velocity, and that the relaxation time varies very weakly over a large energy change.

By defining the energy relaxation length λ_w as

$$\lambda_w = \frac{5}{3} v\tau_{wn}, \quad (12)$$

the solution (11) is found to be

$$w(x) = w_0 + \Delta w(0) \exp(-x/\lambda_w) + \frac{3}{5} \int_0^x qE(\xi) \exp[(\xi - x)/\lambda_w] d\xi, \quad (13)$$

where $\Delta w(x) = w(x) - w_0$. Equation (13) shows that the energy distribution is basically given, apart from the exponentially decaying initial condition, by the convolution integral between the force acting upon the carriers and an exponential kernel having a characteristic length λ_w .

Figure 22 shows the calculated energy distribution $w(x)$ within the collector space-charge region for three different values of λ_w . For the smallest value, the exponential kernel behaves like a delta function and the energy profile approaches the electric field pattern. By contrast, for the largest value of λ_w , the exponential kernel is approximately constant, and the convolution tends to the potential distribution within the collector. For intermediate values, however, neither one of the previous approximations holds.

A possible energy dependence for the electron ionization coefficient α_n can be derived from the Chynoweth expression [35], accounting for the energy-field relationship which holds in static conditions. The result is

$$\alpha_n(w) = A \exp(-E_{th}/\Delta w), \quad (14)$$

in which E_{th} is related to the impact-ionization threshold energy E_{ii} by the expression

$$E_{th} = \frac{3}{5} \frac{\lambda_w}{\lambda_p} E_{ii}, \quad (15)$$

in which λ_p is the optical-phonon-projected mean free path [23].

Here we assume that the electron-ionization coefficient α_n can be expressed as a function of the electron energy w . We realize that such an assumption may be controversial,

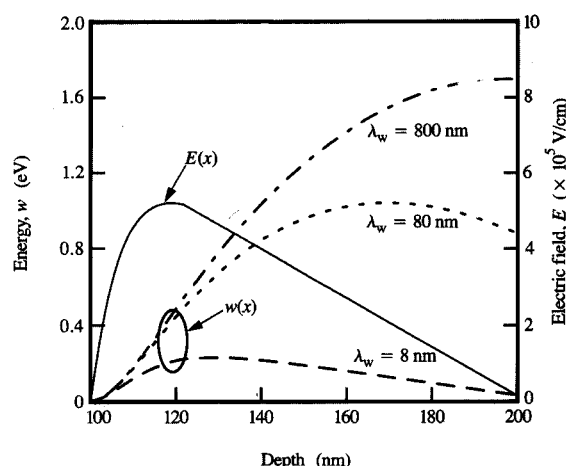


Figure 22

Calculated energy distributions $w(x)$ in the collector space-charge region of a bipolar transistor assuming various values of the energy relaxation length λ_w . The assumed field distribution $E(x)$ is also shown.

since in principle impact ionization is affected by the high-energy tail of the distribution function which, under highly nonstatic conditions, may be loosely related to the mean carrier energy. However, we believe that such an assumption holds in most cases, and that it represents a higher-order approximation compared with the standard field-dependent assumption. Its obvious advantage is that it accounts for nonlocal effects, which manifest themselves as *dead regions* where, despite the large field, no impact ionization can occur because of an insufficient potential drop along the current path.

The multiplication factor $M - 1$ (the ratio of the generated electron-hole pairs to the number of incident carriers) has been calculated by neglecting the hole contribution to impact ionization, i.e.,

$$M - 1 = \exp \left\{ \int_0^{x_d} \alpha_n(w) dx \right\} - 1 \approx \int_0^{x_d} \alpha_n(w) dx. \quad (16)$$

The latter expression is applicable for small values of the integral at the exponent, i.e., for low impact-ionization rates.

As shown in **Figure 23**, Equation (16) fits the experimental data far better than the standard field-dependent Chynoweth formulation does, indicating that, to first order, the temperature dependence of the ionization coefficient captures the most important features of the impact-ionization mechanism.

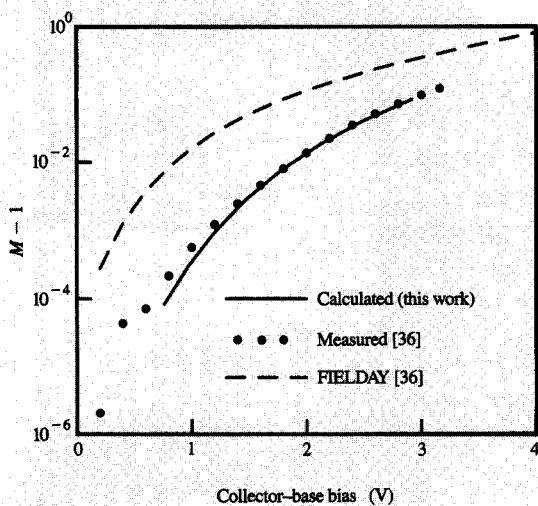


Figure 23

Calculated and measured multiplication factor ($M - 1$) versus collector-base bias at a base-emitter bias of 0.7 V. Conventional electric field models predict too much carrier multiplication.

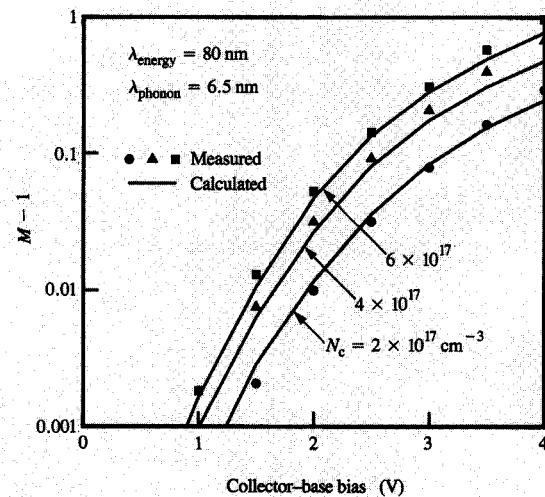


Figure 24

Calculated and measured multiplication factors for various experimental collector doping levels. The same value for λ_w (80 nm) matches these and a variety of other data.

By using λ_w as the *only* fitting parameter, the multiplication factor of a variety of different base and collector profiles can be modeled very accurately, as shown by the curves of **Figure 24**.

Although it is well known that impact ionization is affected by details of the energy tail of the carrier distribution, as described in the Monte Carlo section below, these results show that the average energy contains first-order information regarding the magnitude and nonlocality of ionization events. There is some validity, therefore, in assuming that this simple energy model can be used to extrapolate the breakdown behavior of heavily doped junctions. The associated breakdown voltage can be calculated by assuming that $M - 1 = 1/\beta$. The results of such an analysis as a function of collector doping level N_c are shown in **Figure 25**.

At very high doping levels, the breakdown mechanism may no longer be dominated by impact ionization, but rather by tunneling. However, uniformly abrupt profiles result in unacceptably low breakdown voltages, beyond $N_c \approx 5 \times 10^{17} \text{ cm}^{-3}$. Also shown is the effect of inserting a lightly doped spacer [36], which increases the breakdown voltage by 0.3 V for a 75-nm effective spacer. As shown by the example in **Figure 25**, the effective spacer width for diffused profiles ranges from 50–100 nm; this has been shown to improve breakdown (cf. **Figure 26**) while not degrading speed performance [37]. Since velocity

overshoot has only a small effect on total device speed, the good correlation of breakdown behavior predicted on the basis of the energy model with experimental results implies achievement of the capability to more effectively optimize the breakdown-speed trade-off of the base-collector junction.

4. Applications of Monte Carlo modeling

As discussed in the previous sections, both the DD and hydrodynamic models provide low-order approximate solutions to the Boltzmann transport equation (BTE). In small devices, in which the distances between scattering events are comparable to device dimensions, only a full solution of the BTE will suffice. Here we are not at liberty to impose or assume specific shapes for the distribution function, we may not enjoy the simplifications of the relaxation time approximation, and we must be prepared to take into account extreme off-equilibrium transport. For these reasons, as devices shrink, the DD and hydrodynamic device modeling formulations [26, 38] must break down, for they are only a truncated version of a complete solution to the BTE. Unfortunately, a direct solution to the BTE coupled to a self-consistent electric field pattern (as obtained from the Poisson equation) for any realistic device structure far outstrips current computing capacity. A direct solution to the BTE amounts to solving, as a function of time and for each carrier type

(holes and electrons) a five- or six-dimensional (two or three real space dimensions, and three wave-vector or velocity dimensions) integro-differential equation.

With a dogged desire to preserve as many fundamental aspects as are relevant to the modeling, we have adopted a statistical rather than a deterministic approach, recognizing that such modeling requires large doses of computing resource. Both of these trade-offs are implied in a Monte Carlo solution to the BTE. We have mitigated the first trade-off by enhancing rare events to allow resolution of the "tail," or highest-energy portion, of the distribution function [39, 40], a procedure unnecessary in a deterministic approach. The second trade-off puts us squarely in the realm of "large-scale computing" and precludes the interactive or short-turnaround calculations prevalent in engineering design cycles. With these goals and trade-offs in mind, the DAMOCLES (Device Analysis using Monte Carlo et Poisson solver) program has been developed over the last five years [41-53]. Currently running on both large IBM mainframes and high-end IBM workstations, it has been used to simulate two-dimensional device structures fabricated using both silicon and compound semiconductor materials. Devices simulated include silicon MOSFET and bipolar devices, and HEMT, MESFET, and bipolar devices fabricated using various III-V compounds and semiconductor alloys. Typically, 10 to 100 IBM 3090S CPU hours are required to complete the solution for a bias point.

In the following section, we illustrate two different applications of the DAMOCLES approach. The first is a comparative study of exploratory MOSFETs fabricated using different materials systems, in an attempt to understand the impact of semiconductor materials on the performance of short-channel MOSFETs [53]. To our knowledge, no other device-simulation approach currently has both the physical rigor and flexibility of implementation to facilitate this study. The second involves a study of profile design for advanced bipolar transistors, in which nonequilibrium and nonlocal effects have been modeled using the DAMOCLES approach.

• Monte Carlo modeling of exploratory MOSFETs

One area in which modeling can be of great assistance is in comparing different device families and different associated materials systems. Each device family or materials system is generally associated with a specific forte: e.g., silicon MOSFETs excel in dense circuits, silicon bipolar devices excel in switching speed and drive, and III-V materials systems offer the highest speeds but limited density. These associations, however, are born of experience at current device dimensions: for example, $\leq 1\text{-}\mu\text{m}$ channel lengths for MOSFETs and $\leq 100\text{-nm}$ base widths for bipolar transistors. We opted to take a very different approach to this problem of comparing different types of devices and

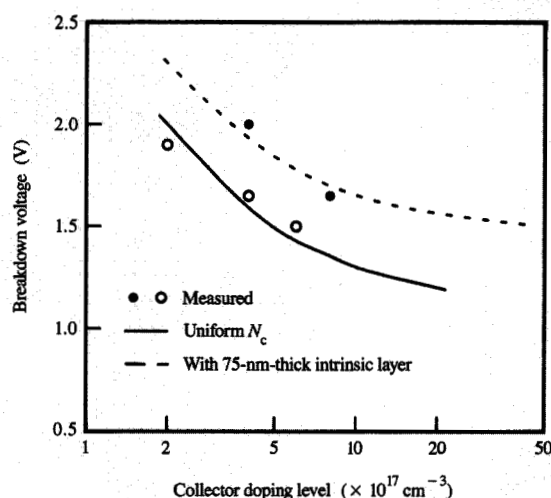


Figure 25

Breakdown voltage at which base current reverses as a function of collector doping.

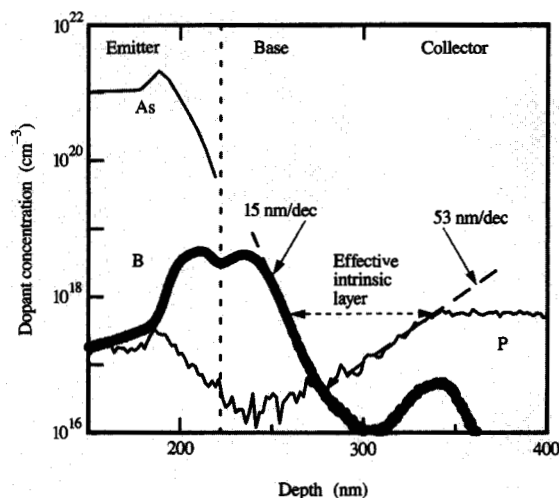


Figure 26

SIMS profile of devices having lightly doped spacers, as described in [37]. The effective spacer width is easily determined using an exponential fit to the doping profile.

materials systems: namely, for extremely small devices ($10\times$ smaller than the above dimensions), inquiring which materials systems would offer the fastest logic performance

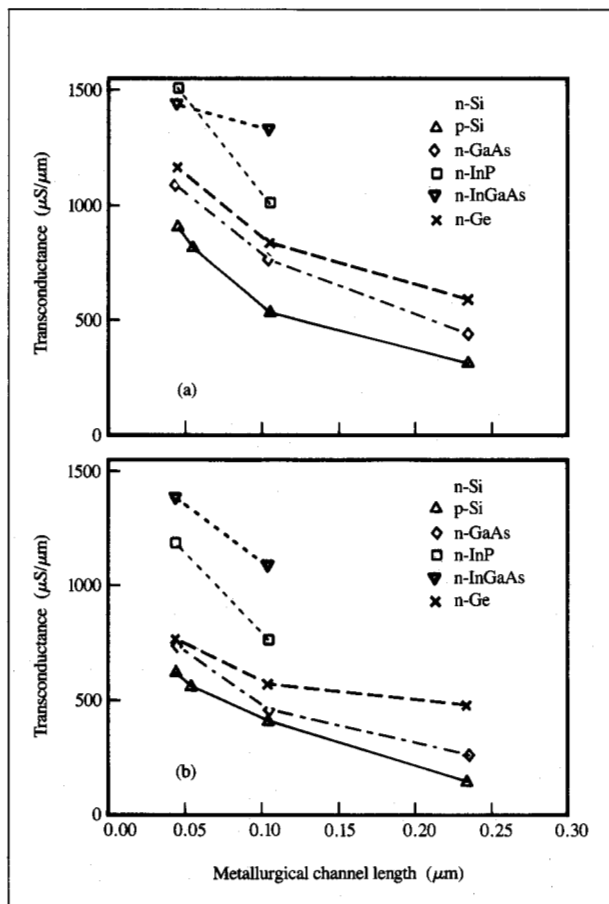


Figure 27

Calculated MOSFET transconductance vs. metallurgical channel length, for five different semiconductor materials at (a) 77 K and (b) 300 K. The results are for electrons in n-MOSFETs except for those depicted in green, which are for holes in silicon p-MOSFETs.

for the *same* device structure. To attempt to address this question, our model needed to blend physical rigor (to be reasonably predictive) and flexibility (to represent different materials systems). Previously, DAMOCLES had demonstrated an excellent capability for predicting terminal currents and transconductances for 0.1- μm silicon n-MOSFETs [42]. For this reason, and because MOSFETs seem to be essential ingredients in all notions of the future of microelectronics and ULSI (ultralarge-scale integration), we adopted the n-MOSFET cross section as our “universal” device structure for materials comparisons. Note that our decision to use only one device structure is not without criticism: Different device structures may (inadvertently) “favor” certain materials systems by precluding optimum performance from a given system. Yet, in defense of our approach, we reaped the benefits of

a “clean” comparison, being able to attribute differences solely to the different systems considered.

In addition to adopting a single device structure for investigation, we made another important assumption: Our metric for device performance was device transconductance. The reasoning behind this decision was as follows: From present trends in ULSI, it is apparent that wiring capacitance is the primary switching load. Assuming that the intrinsic delay of the transistor switch is much less than the loaded switching time, the current flowing during switching looks essentially constant. Accordingly, switching time should be determined by how much charge a transistor can deliver to or remove from the capacitive load according to $\Delta t = \Delta Q/I = C\Delta V/I$, where a constant current I is assumed to be acting on a capacitive load C over a time interval Δt , with voltage and charge changes ΔV and ΔQ , respectively. This can be simplified to $\Delta t = C/g_m$, where g_m is the transconductance $\Delta I/\Delta V$ and $\Delta I = I$; i.e., g_m is a large-signal transconductance. However, in short-channel MOSFETs with significant velocity saturation effects, the transconductance is approximately constant with device bias, permitting the small-signal and large-signal transconductance to be used interchangeably. Finally, we infer that switching speed, proportional to $(\Delta t)^{-1}$, is given by g_m/C , where C is the wiring capacitance.

For every device simulated, the same doping profile was used in the channel regions; however, in the heavily doped source and drain regions, doping in the GaAs, InP, and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ devices considered was assumed to be an order of magnitude lower than in the Si and Ge devices considered. This decision was made because 1) donor concentrations $\geq 10^{20} \text{ cm}^{-3}$ are unrealistic in compound semiconductors, and 2) such high doping, even if possible, invites the possibility of electrons occupying the L- and X-valleys *prior* to entering the channel, which is detrimental to device performance.

The bias at which device performance is determined depends on the MOSFET channel length in order to keep the electric field strength approximately constant. For a 0.25- μm channel length, drain and gate biases range up to 2.5 V; for 0.10- and 0.07- μm channel lengths, this is decreased to 1.0 and 0.7 V, respectively.

The results of our modeling are shown in **Figure 27**. There, device transconductance is plotted versus metallurgical channel length at two temperatures (77 and 300 K) and for five semiconductors: Si, GaAs, InP, $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, and Ge. Electrons are the carriers that carry the current in the MOSFET channel (i.e., these are n-MOSFETs) except in the case labeled “p-Si,” which represents our results for hole channels in silicon p-MOSFETs. A few interesting trends are discernible. First, focusing only on the Si devices, our calculations indicate that the performance of the p-MOSFET should

improve markedly relative to that of the n-MOSFET as channel lengths decrease from ~ 0.25 to $0.05 \mu\text{m}$. At $0.25 \mu\text{m}$, the n-MOSFET has twice the transconductance of the p-MOSFET; at $0.05 \mu\text{m}$, this advantage is reduced to about 20%. This relative improvement in the expected performance of the p-MOSFET comes about because of the greater increase in the channel velocity in the p-MOSFET as the dimensions shrink. This is consistent with Figure 28, which shows the carrier velocity (electrons in n-MOSFETs and holes in p-MOSFETs) versus the normalized distance along the channel for the Si devices at the largest ($0.233\text{-}\mu\text{m}$) and smallest ($0.043\text{-}\mu\text{m}$) channel lengths considered, at 300 K. Figure 28 clearly shows that only for the last $\sim 15\%$ distance along the channel do velocities exceed the saturated velocity; the remainder of the channel is in quasi-equilibrium with the local electric field. This is quite different at the smallest channel length, where both electrons and holes are expected to experience strong off-equilibrium effects, as indicated by the broad velocity maximum.

The second indication from Figure 27 is that electrons in GaAs seem to offer little advantage over those in Si or Ge (and even holes in Si) as channel dimensions shrink to $\leq 0.1 \mu\text{m}$. This certainly differs from the preconceived notions of the materials advantages mentioned earlier in this section. Yet, Figure 27 also shows two impressive performers, namely InP and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ MOSFETs. This rather surprising result can be explained by three effects: loss of gate capacitance, loss of velocity overshoot, and carrier heating.

The transconductance of a MOSFET is proportional to the capacitance between its gate and channel, for it is this coupling which permits the channel charge to be modulated by the gate voltage. This capacitance constitutes a series combination of geometric capacitance involving insulator thickness and permittivity, and a capacitance related to the position of the charge centroid below the insulator-semiconductor interface. In materials having a low density of states (DOS), i.e., low effective mass, the charge centroid resides farther below the interface than for high-effective-mass materials. This occurs because a low-DOS material requires a large change in Fermi energy to accommodate a given change in inversion charge, which indicates a reduced capacitive coupling between surface potential and inversion charge. Thus, even for the invariant device geometry, the silicon devices have the highest effective gate capacitance.

In small devices where strong off-equilibrium transport effects are present, a second explanation for the similar transconductance among different materials can be found, namely loss of velocity overshoot. Transconductance measures how much the drain current increases for a given increase in gate voltage. Two separate increases determine the total current increase: the increase in the channel

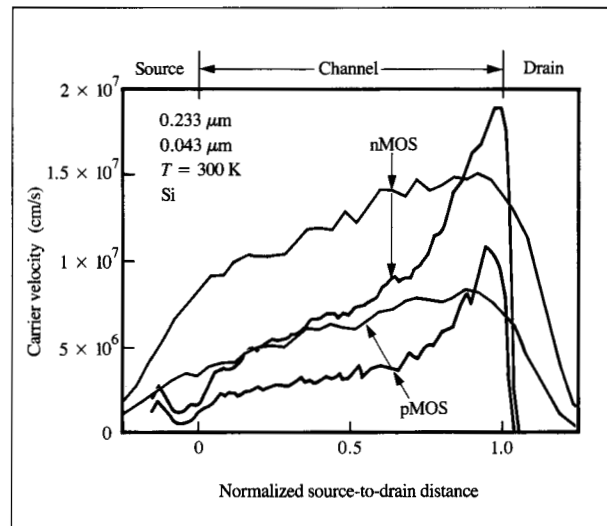


Figure 28

Calculated carrier velocity versus normalized distance from source to drain for silicon n- and p-MOSFETs at 300 K. The curves apply to two metallurgical channel lengths: $0.233 \mu\text{m}$ (red) and $0.043 \mu\text{m}$ (green).

charge and the increase in the channel velocity. The previous paragraph discussed the first component, i.e., the capacitive coupling between gate and channel charge. The increase in the channel velocity as gate voltage increases is dependent on whether the channel is already in a strong overshoot regime or in a quasi-equilibrium regime. For the former case, the current is already enhanced because of velocity overshoot prior to the increase in gate bias. Simply increasing the gate bias and expecting a proportional gain in current because of an increase in velocity overshoot only applies up to a point, beyond which the benefits of velocity overshoot are exhausted. From a terminal current perspective, this appears as a loss of velocity overshoot. However, if the channel is in quasi-equilibrium initially, increasing the gate bias renders the drain-to-source field in the channel more uniform, i.e., raises the drain-to-source field near the source. This larger driving force pushes the carriers toward and into the velocity overshoot regime, reaping relatively higher increases in drain current and transconductance.

Finally, carrier heating is responsible for the similar transconductance among the different materials systems. High carrier kinetic energies occur both from the applied bias and from the Pauli exclusion principle. The former effect is obvious: When scattering becomes relatively less frequent because of shrinking device geometry, carriers are free to probe the semiconductor band structure away from local band minima up to energies of the order of the device bias. Effective mass and mobility, properties associated

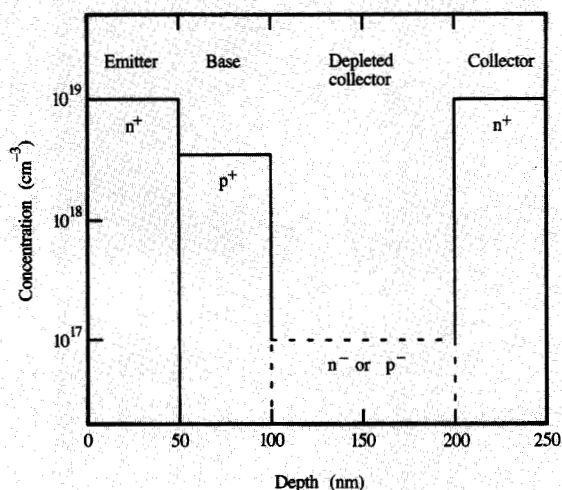


Figure 29

Assumed doping profiles of the depleted n- and p-type collector designs under consideration.

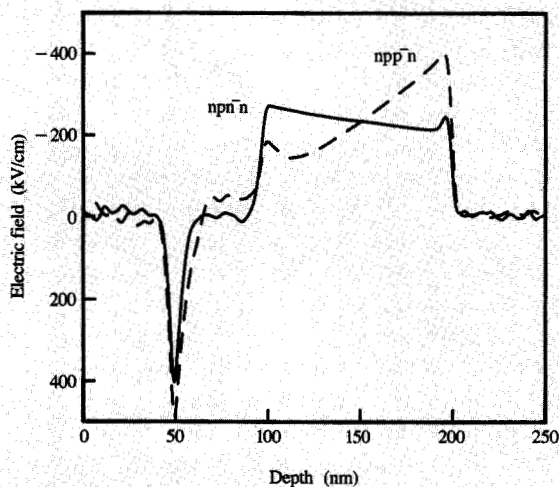


Figure 30

Calculated electric field profiles under operating conditions, assuming the collector designs of Figure 29. Note that the profiles peak at opposite sides of the junction for the n- and p-type cases.

in the Brillouin zone. The Pauli principle also comes into play: Small devices are almost always heavily doped devices (to obtain the necessary electric field confinement). Furthermore, higher densities of mobile carriers are necessary if small devices are to deliver ever-higher current densities. These high carrier densities, together with the Pauli principle, serve to move the carrier states away from the valley minima. In GaAs, there is a severe penalty for this: The L- and X-valleys are associated with states having a high effective mass and low group velocity. Occupancy of these states depresses GaAs device performance.

For the InP and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ MOSFETs, the Γ -L valley separation is much greater (~ 0.72 and 0.83 eV, respectively), permitting the low effective mass transport of the Γ valley to dominate throughout the cases considered here. Only at 77 K and at the smallest channel length do we see evidence of the different Γ -L separation in these two types of devices, personified by the crossover in their transconductance in Figure 27(a).

Are these conclusions about the relative performance of devices fabricated from the different materials generally valid? In a transport regime in which scattering is relatively infrequent (but *not* entirely absent), scattering deposits carriers into states with the highest DOS. These states are characterized by the lowest group velocities, for the DOS is just a weighted average over the inverse group velocity. Thus, we propose that the DOS can be used as a measure of the anticipated (inverse) velocities to be found in materials in the weak scattering regime. For Si (conduction and valence bands), Ge, and GaAs, the DOS are remarkable similar at around 1 eV, namely $10^{22} \text{ cm}^{-3} \text{ eV}^{-1}$, while for $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ and InP, the DOS is 10 – $50\times$ smaller at similar energies. This quantitative difference is borne out by our calculations.

• Speed and breakdown trade-offs for silicon bipolar transistors

In bipolar transistors having basewidths less than 50 nm, electron transport in the base-collector junction is of great importance for two reasons: 1) the time constant associated with signal propagation across the space-charge region, the base-collector signal delay, becomes an appreciable component of the total transit time [32, 46]; 2) the high electric field present in that region leads to impact ionization, imposing a serious constraint on vertical scaling. If the electric field changes rapidly over distances comparable to the electron mean free path, velocity overshoot occurs. A DD formalism does not adequately address such nonequilibrium transport, as discussed in Section 3. Hydrodynamic modeling is useful as a design tool and can be used to predict with reasonable accuracy currents generated by impact ionization. However, it is ultimately based on an incorrect physical basis because it

with the valley minima, become increasingly irrelevant. Rather, group velocity and the DOS increasingly dictate how fast carriers will move and where they will be found

requires an *a priori* assumption about the carrier distribution. Monte Carlo modeling does not require such an assumption and, as such, is ideally suited for the study of impact ionization and velocity overshoot. As an illustration, two idealized silicon bipolar transistor profiles are compared in this section, with particular attention to nonlocal phenomena in the base-collector region.

The profiles under consideration are shown in **Figure 29** and consist of a 50-nm-wide base and a 100-nm lightly doped depleted region, either n- or p-type. The depleted n- and p-type base-collector designs represent two extreme cases in which the maximum electric field for zero collector current is either located near the base or near the heavily doped subcollector.

The calculated electric field profiles under operating conditions ($V_{BE} = 0.75$ V and $V_{CB} = 1.5$ V, $J_C \approx 10^5$ A/cm²) are shown in **Figure 30**. The high collector current tends to flatten the electric field profile of the n-type design and increase the maximum field of the p-type design. The average electron energy peaks deeply within the space-charge region, as illustrated in **Figure 31**. In the case of the extended-base profile, the electron energy profile follows the electric field profile. Associated transport parameters such as average electron velocity and impact ionization coefficient are expected to be well approximated by a local model, i.e., one that depends on the electric field at every point. For the depleted n⁻ collector design, the electric field and average electron energy were found to peak at the two opposite ends of the base-collector space-charge region, suggesting strong nonlocal behavior.

DAMOCLES uses the Keldysh formula with a 1.2-eV isotropic energy threshold to derive the rate of impact ionization as a function of carrier energy [41]. The ionization coefficient is then calculated at every point of the collector as the ratio of the probability of impact ionization per unit time to the average electron velocity. Results of the calculation are shown in **Figure 32**. The peak in the ionization coefficient for the extended-base design is twice that of the n⁻ design, reflecting the higher average energy of the electrons at that location (**Figure 31**).

A conventional local model expressing the ionization coefficient as a function of the electric field at every point should model reasonably well the extended-base profile but not the n⁻ collector design. It is interesting to note that neither a drift-diffusion model nor a hydrodynamic model is able to predict impact ionization events occurring in the heavily doped subcollector because it is a field-free region. **Figure 31** indicates that the ionization coefficient is not zero, but quite small there for both profiles.

The calculated electron velocity profiles are markedly different for the two designs in the base-collector space-charge region, as shown in **Figure 33**. The velocity is very low in the emitter and the subcollector, and it increases

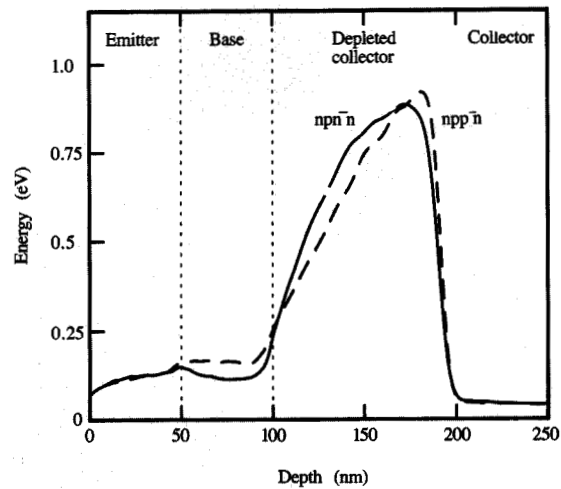


Figure 31

Calculated electron energy profiles, assuming the collector designs of **Figure 29**. The profiles peak deeply within the space-charge region; the higher maximum is reached by the profile for the depleted p-type design.

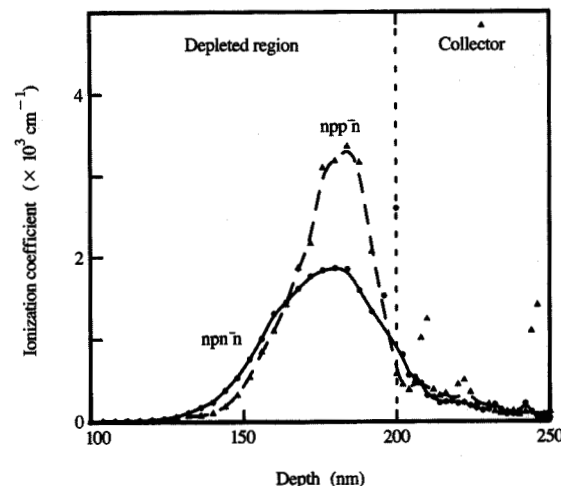


Figure 32

Calculated ionization coefficient for the profiles of **Figure 29**.

steadily as the electrons traverse the base toward the collector. It peaks at a value of 1.8×10^7 cm/s for the n⁻ collector design, and the peak occurs at the base edge

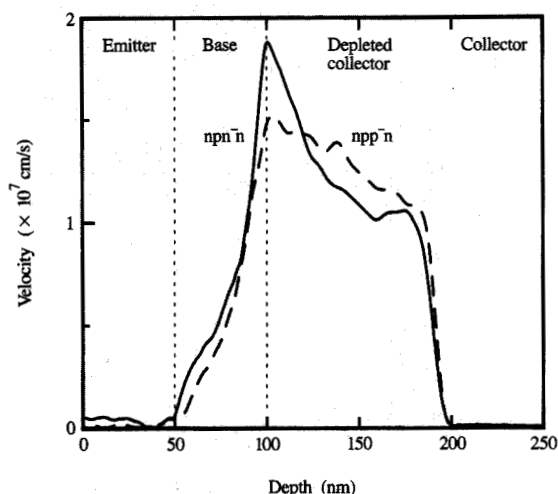


Figure 33

Calculated velocity profiles for the n- and p-type base-collector designs.

of the space-charge region. This velocity overshoot is expected whenever the electric field varies significantly over distances comparable to an electron mean free path, and the energy is still small. The electric field does indeed rise very sharply at the location of the velocity overshoot. However, the velocity drops quite rapidly, and the region of significant velocity overshoot is no wider than 40 nm. For the extended-base profile, the electric field rises smoothly in the base-collector region and, as a result, the velocity does not overshoot significantly. For both profiles, the velocity is not much higher than the saturation velocity over more than 50% of the base-collector space-charge region.

The impact of velocity overshoot on the speed of the transistor can be assessed by calculating the base-collector signal delay following the analysis in [54] and comparing it to the case in which no velocity overshoot is present. The calculated delays are 0.41 and 0.46 ps for the n^-n^+ and p^-n^+ designs, respectively, compared to 0.52 ps for $v = v_{\text{sat}}$. This 21% best-case decrease in signal delay translates into a 6% reduction of the overall transit time of a silicon transistor having an 80-GHz cutoff frequency, for which the base transit time dominates the total transit time. In sharp contrast to III-V heterojunction bipolar transistors, velocity overshoot in the base-collector region of silicon bipolar transistors is therefore expected to have little influence on their high-frequency performance.

These two examples have illustrated some of the capabilities of Monte Carlo simulation for modeling transport phenomena that cannot be suitably modeled by drift-diffusion or hydrodynamic formalisms. They illustrate the use of an approach that can be used to obtain an improved understanding of nonequilibrium transport aspects, and they point to a need for a method that permits the simulation of several profiles in a relatively short time, each profile at several bias points. This is not yet possible with DAMOCLES.

Conclusions and directions for future work

Several examples of simulations of advanced MOSFETs and bipolar transistors have been presented here to illustrate the need for more physical modeling than is currently used in order to design such devices. Emphasis has been on the numerical modeling of three-dimensional, nonequilibrium, and nonlocal phenomena which are expected to play increasing roles in the further design of semiconductor devices used in digital applications. However, this is not to say that numerical models are expected to replace intuition. On the contrary, even better intuition and imagination will be required for setting up the relevant structures for simulations, using the proper and appropriate models for the problems at hand, as well as digesting and interpreting the vast amount of information resulting from the simulations. Moreover, an understanding of solid-state physics will be necessary in order to understand the implications and limitations of the physical models incorporated in the simulation tools.

Much work is still needed to improve the physical understanding as well as the numerical implementation of both the hydrodynamic and the Monte Carlo models. The models discussed in this paper provide either an approximate or an exact solution to the semiclassical Boltzmann transport equation. Because of its semiclassical nature, that equation cannot by itself account for many quantum phenomena commonly encountered in state-of-the-art semiconductor devices, such as tunneling either between heavily doped semiconductor junctions or through thin insulators, and quantization effects in the two-dimensional electron gas formed at the Si-SiO₂ interface in MOSFETs. To model these quantum-mechanical effects, either a phenomenological approach must be incorporated into the framework of the Boltzmann transport equation or a complete quantum-mechanical description of carrier transport [55, 56] must be used.

Other important areas requiring significant attention include the representation of three-dimensional device structures and visualization methods. Efficient and flexible representation and discretization of arbitrary geometrical device structures composed of different materials are needed for relevant 3D device simulations. Of equal importance is the improvement in visualization methods.

Graphical representation of voluminous output data generated from 3D simulations remains a challenge [57]. Fast visualization tools will be needed in order to make it possible to examine and interpret simulation results in a timely manner.

Finally, it is our belief that the continuing increase in the cost of semiconductor fabrication facilities and the decrease in the cost of computation should make the numerical modeling of semiconductor devices an increasingly important part of microelectronic technology development.

Acknowledgments

One of the authors (W. Lee) wishes to thank members of the Technology Simulation Group at IBM Technology Products, Burlington, Vermont, for their FIELDAY support. He is especially grateful to T. Linton for AIX® FIELDAY support and to the late B. Hartigan for providing the translator software for related 3D graphics aspects. Another of the authors (J. Mandelman) wishes to thank T. Way for his help in obtaining the color renditions of the FIELDAY output.

3090 and RISC System/6000 are trademarks, and AIX is a registered trademark, of International Business Machines Corporation.

References

1. E. M. Buturla, P. E. Cottrell, B. M. Grossman, and K. A. Salsburg, "Finite-Element Analysis of Semiconductor Devices: The FIELDAY Program," *IBM J. Res. Develop* **25**, 218 (1981).
2. W. L. Engl, H. K. Dirks, and B. Meinerzhagen, "Device Modeling," *Proc. IEEE* **71**, 10 (1983).
3. S. P. Gaur, P. A. Habitz, Y.-J. Park, R. K. Cook, Y.-S. Huang, and L. F. Wagner, "Two-Dimensional Device Simulation Program: 2DP," *IBM J. Res. Develop* **29**, 242 (1985).
4. D. C. Cole, E. M. Buturla, S. S. Furkay, K. Varahramyan, J. Slinkman, J. A. Mandelman, D. P. Foty, O. Bula, A. W. Strong, J. W. Park, T. D. Linton, J. B. Johnson, M. V. Fischetti, S. E. Laux, P. E. Cottrell, H. G. Lustig, F. Pileggi, and D. Katcoff, "The Use of Simulation in Semiconductor Technology Development," *Solid-State Electron.* **33**, 591 (1990).
5. S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, Vienna, 1984.
6. C. M. Snowden, *Introduction to Semiconductor Device Modeling*, World Scientific Press, Singapore, 1986.
7. Carlo Jacoboni and Paolo Lugli, *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag, Vienna, 1989.
8. D. Kenney, E. Adler, B. Davari, J. DeBrosse, W. Frey, T. Furukawa, P. Geiss, D. Harmon, D. Horak, M. Kerbaugh, C. Koburger, J. Lasky, J. Rembetski, W. Schwitek, and E. Sprogis, "16-Mbit Merged Isolation and Node Trench SPT Cell (MINT)," *Technical Digest of the 1988 VLSI Symposium on VLSI Technology*, San Diego, 1988, pp. 25-26.
9. S. F. Geissler and J. A. Mandelman, "Analysis of Sidewall Lateral Parasitic Leakage in a 16-Mb DRAM Cell," presented at the Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits, NUPAD III, Honolulu, Hawaii, 1990.
10. B. Davari, C. Koburger, T. Furukawa, Y. Taur, W. Noble, J. Wamock, and J. Mauer, "A Variable Stress Shallow Trench Isolation (STI) Technology with Diffused Sidewall Doping for Submicron CMOS," *Technical Digest of the International Electron Devices Meeting*, 1988, p. 92.
11. J. A. Appels, E. Kooi, M. M. Paffen, J. J. H. Schatorje, and W. H. C. G. Verkuyl, "Local Oxidation of Silicon and Its Application in Semiconductor-Device Technology," *Philips Res. Repts.* **25**, 118 (1970).
12. E. M. Buturla, J. Johnson, S. Furkay, and P. Cottrell, "A New Three-Dimensional Device Simulation Formulation," *NASECODE VI: Proceedings of the Sixth International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits*, J. J. H. Miller, Ed., Boole Press Ltd., Dublin, 1989, p. 291.
13. J. Burghartz, J. Sun, S. Mader, C. Stanis, and B. Ginsberg, "Perimeter and Plug Effects in Deep Sub-micron Polysilicon Emitter Bipolar Transistors," *Technical Digest of the 1990 Symposium on VLSI Technology*, 1990, p. 55.
14. S. Konaka, T. Kobayashi, T. Matsuda, M. Ugajin, K. Imai, and T. Sakai, "HSST/BICMOS Technology with 26 ps ECL and 45 ps 2V CMOS Inverter," *Technical Digest of the International Electron Devices Meeting*, 1990, p. 493.
15. A. Yoshii, H. Kitazawa, M. Tomizawa, S. Horiguchi, and T. Sudo, "A Three-Dimensional Analysis of Semiconductor Devices," *IEEE Trans. Electron Dev.* **ED-29**, 184 (1982).
16. P. Conti, G. Heiser, and W. Fichtner, "Three-Dimensional Transient Simulation of Complex Silicon Devices," *Jpn. J. Appl. Phys.* **29**, L2271 (1990).
17. C. T. Kirk, "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities," *IRE Trans. Electron Dev.* **ED-9**, 164 (1962).
18. K. Blotekjaer, "High-Frequency Conductivity, Carrier Waves, and Acoustic Amplification in Drifted Semiconductor Plasmas," *Ericsson Technics* **22**, 125 (1966).
19. A. Gnudi, F. Odeh, and M. Rudan, "Investigation of Non-Local Transport Phenomena in Small Semiconductor Devices," *European Trans. Telecommunications & Related Technologies* **1**, 307 (1990).
20. M. Rudan and F. Odeh, "Multi-Dimensional Discretization Scheme for the Hydrodynamic Model of Semiconductor Equations," *COMPEL* **5**, 149 (1986).
21. K. Katayama and T. Toyabe, "A New Hot Carrier Simulation Method Based on Full 3D Hydrodynamic Equations," *Technical Digest of the International Electron Devices Meeting*, 1989, p. 135.
22. B. Meinerzhagen, "Consistent Gate and Substrate Current Modeling Based on Energy Transport and the Lucky Electron Concept," *Technical Digest of the International Electron Devices Meeting*, 1988, p. 504.
23. W. Shockley, "Problems Related to p-n Junctions in Silicon," *Solid-State Electron.* **2**, 35 (1961).
24. E. Sholl and W. Quade, "Effect of Impact Ionization on Hot-Carrier Energy and Momentum Relaxation in Semiconductors," *J. Phys. C* **20**, L861 (1987).
25. B. Davari, W. Chang, C. Oh, Y. Taur, M. Wordeman, D. Moy, K. Petrillo, J. Bucchignano, H. Ng, M. Rosenfield, F. Horn, and M. Rodriguez, "A High-Performance 0.25- μ m CMOS Technology," *Technical Digest of the International Electron Devices Meeting*, 1988, p. 56.
26. A. Forghieri, R. Guerrieri, P. Ciampolini, A. Gnudi, M. Rudan, and G. Baccarani, "A New Discretization Strategy of the Semiconductor Equations Comprising Momentum and Energy Balance," *IEEE Trans. Computer-Aided Design* **7**, 231 (1988).
27. W. Haensch and M. Miura-Mattausch, "The Hot-Electron Problem in Small Semiconductor Devices," *J. Appl. Phys.* **60**, 650 (1986).

28. G. Baccarani and M. R. Wordeman, "Transconductance Degradation in Thin-Oxide MOSFET's," *IEEE Trans. Electron Dev.* **ED-30**, 1295 (1983).
29. J. A. Cooper and D. F. Nelson, "High-Field Drift Velocity of Electrons at the Si-SiO₂ Interface as Determined by a Time-of-Flight Technique," *J. Appl. Phys.* **54**, 1445 (1983).
30. R. v. Overstraeten and H. De Man, "Measurements of the Ionization Rates in Diffused Silicon p-n Junctions," *Solid-State Electron.* **13**, 583 (1970).
31. S. S. Iyer, G. L. Patton, J. M. C. Stork, B. S. Meyerson, and D. L. Harame, "Heterojunction Bipolar Transistors Using Si-Ge Alloys," *IEEE Trans. Electron Dev.* **36**, 2043 (1989).
32. M. Ugajin, S. Konaka, K. Yokoyama, and Y. Amemiya, "A Simulation of High-Speed Silicon Heteroemitter Bipolar Transistors," *IEEE Trans. Electron Dev.* **36**, 1102 (1989).
33. P. F. Lu and T. C. Chen, "Collector-Base Junction Avalanche Effects in Advanced Double-Poly Self-Aligned Bipolar Transistors," *IEEE Trans. Electron Dev.* **36**, 1182 (1989).
34. G. Baccarani and M. R. Wordeman, "An Investigation on Steady-State Velocity Overshoot in Silicon," *Solid-State Electron.* **28**, 407 (1985).
35. A. G. Chynoweth, "Ionization Rates for Electrons and Holes in Silicon," *Phys. Rev.* **109**, 1537 (1958).
36. D. D. Tang and P.-F. Lu, "A Reduced-Field Design Concept for High-Performance Bipolar Transistors," *IEEE Electron Device Lett.* **10**, 67 (1989).
37. J. Comfort, E. F. Crabbé, G. L. Patton, J. M. C. Stork, J. Y.-C. Sun, and B. S. Meyerson, "Impact Ionization Reduction in SiGe HBTs," *Abstracts of the Device Research Conference*, 1990, Paper VIA-3.
38. B. Meinertzhagen and W. L. Engl, "The Influence of the Thermal Equilibrium Approximation on the Accuracy of Classical Two-Dimensional Numerical Modeling of Silicon Submicron MOS Transistors," *IEEE Trans. Electron Dev.* **35**, 689 (1988).
39. A. Phillips, Jr. and P. J. Price, "Monte Carlo Calculations on Hot Electron Energy Tails," *Appl. Phys. Lett.* **30**, 528 (1977).
40. E. Sangiorgi, B. Riccò, and F. Venturi, "MOS²: An Efficient Monte Carlo Simulator for MOS Devices," *IEEE Trans. Computer-Aided Design* **7**, 259 (1988).
41. M. V. Fischetti and S. E. Laux, "Monte Carlo Analysis of Electron Transport in Small Semiconductor Devices Including Band-Structure and Space-Charge Effects," *Phys. Rev. B* **138**, 9721 (1988).
42. S. E. Laux and M. V. Fischetti, "Monte Carlo Simulation of Submicrometer Si n-MOSFETs at 77 and 300 K," *IEEE Electron Device Lett.* **EDL-9**, 467 (1988).
43. M. V. Fischetti and S. E. Laux, "Monte Carlo Simulation of Submicron Si MOSFETs," *Simulation of Semiconductor Devices and Processes*, Vol. 3, G. Baccarani and M. Rudan, Eds., Technoprint, Bologna, Italy, 1988, p. 349.
44. M. V. Fischetti, S. E. Laux, and D. J. DiMaria, "The Physics of Hot Electron Degradation of Si MOSFETs: Can We Understand It?" *Appl. Surf. Sci.* **39**, 578 (1989).
45. M. V. Fischetti, S. E. Laux, and W. Lee, "Monte Carlo Simulation of Hot-Carrier Transport in Real Semiconductor Devices," *Solid-State Electron.* **32**, 1723 (1989).
46. W. Lee, S. E. Laux, M. V. Fischetti, and D. D. Tang, "Monte Carlo Simulation of Non-Equilibrium Transport in Ultra-Thin Base Si Bipolar Transistors," *Technical Digest of the International Electron Devices Meeting*, 1989, p. 473.
47. M. V. Fischetti and S. E. Laux, "Are GaAs MOSFETs Worth Building? A Model-Based Comparison of Si and GaAs n-MOSFETs," *Technical Digest of the International Electron Devices Meeting*, 1989, p. 481.
48. S. E. Laux, M. V. Fischetti, and D. J. Frank, "Monte Carlo Analysis of Semiconductor Devices: The DAMOCLES Program," *IBM J. Res. Develop.* **34**, 466 (1990).
49. S. Tiwari, M. V. Fischetti, and S. E. Laux, "Transient and Steady-State Overshoot in GaAs, InP, Ga_{0.47}In_{0.53}As and InAs Bipolar Transistors," *Technical Digest of the International Electron Devices Meeting*, 1990, p. 435.
50. E. F. Crabbé, J. M. C. Stork, G. Baccarani, M. V. Fischetti, and S. E. Laux, "The Impact of Non-Equilibrium Transport on Breakdown and Transit Time in Bipolar Transistors," *Technical Digest of the International Electron Devices Meeting*, 1990, p. 463.
51. S. E. Laux and M. V. Fischetti, "The DAMOCLES Monte Carlo Device Simulation Program," *Computational Electronics: Semiconductor Transport and Device Simulation*, K. Hess, J. P. Leburton, and U. Ravaioli, Eds., Kluwer Academic Publishers, Boston, 1991, p. 87.
52. M. V. Fischetti, "Monte Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zinc-Blende Structures. Part I: Homogeneous Transport," *IEEE Trans. Electron Dev.* **38**, 634 (1991).
53. M. V. Fischetti and S. E. Laux, "Monte Carlo Simulation of Transport in Technologically Significant Semiconductors of the Diamond and Zinc-Blende Structures. Part II: Submicron MOSFETs," *IEEE Trans. Electron Dev.* **38**, 650 (1991).
54. S. E. Laux and W. Lee, "Collector Signal Delay in the Presence of Velocity Overshoot," *IEEE Electron Device Lett.* **11**, 174 (1990).
55. W. R. Frensley, "Quantum Transport Modeling of Resonant-Tunneling Devices," *Solid-State Electron.* **31**, 739 (1988).
56. B. A. Mason and K. Hess, "Quantum Monte Carlo Calculations of Electron Dynamics in Dissipative Solid-State Systems Using Real-Time Path Integrals," *Phys. Rev. B* **39**, 5051 (1989).
57. E. J. Farrell, S. E. Laux, P. L. Corson, and E. M. Buturla, "Animation and 3D Color Display of Multiple-Variable Data: Application to Semiconductor Design," *IBM J. Res. Develop.* **29**, 302 (1985).

Received June 20, 1991; accepted for publication April 28, 1992

Wai Lee *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (WAILEE at WATSON, wailee@watson.ibm.com).* Dr. Lee received an S.B. degree in 1983, an S.M. degree in 1986, and a Ph.D. degree in 1988, all in electrical engineering, from the Massachusetts Institute of Technology. In 1983, he was employed at the Thomas J. Watson Research Center, working on self-aligned GaAs MESFET and HEMT devices. Since July 1988, he has been at the Research Center as a Research Staff Member, initially to work on the simulation of small silicon bipolar devices and the design of high-performance bipolar devices and circuits. He is currently working on high-performance CMOS circuit design. Dr. Lee is a member of the Institute of Electrical and Electronics Engineers.

Steven E. Laux *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (LAUX at WATSON, laux@watson.ibm.com).* Dr. Laux received a B.S.E.E. degree from Purdue University in 1976 and M.S.E. and Ph.D. degrees in electrical engineering from the University of Michigan in 1977 and 1981, respectively. He has since been employed at the Thomas J. Watson Research Center, working first in the Silicon Devices and Technology group, and more recently in the Semiconductor Physics and Devices group. In 1989, he received an IBM Outstanding Innovation Award for the creation of the DAMOCLES program. His interests include device physics and modeling, large-scale computing, and applied mathematics. Dr. Laux is a Senior Member of the Institute of Electrical and Electronics Engineers.

Massimo V. Fischetti *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (FISCHET at WATSON, fischet@watson.ibm.com).* Dr. Fischetti graduated from the University of Milan, Italy, in 1974 with a "Laurea" in physics, and received a Ph.D. in physics from the University of California in 1978. He joined the Thomas J. Watson Research Center in 1983, after four years of experience in experimental solid-state physics at the Physics Laboratories of SGS-Thomson and 3M-Italy. Dr. Fischetti has worked on experimental and theoretical aspects of the degradation of thin silicon dioxide films and on the theory of electron transport in insulators, receiving one IBM Outstanding Innovation Award for the Monte Carlo simulation of electron transport in SiO₂ in 1986, and another for the creation of the DAMOCLES program in 1989. Since 1985, he has been working on the Monte Carlo modeling of electron transport in small semiconductor devices.

Giorgio Baccarani *Dipartimento di Elettronica, University of Bologna, viale Risorgimento 2, 40136 Bologna, Italy (baccarani%bodem7.cineca.it at iceneca.cineca.it).* Professor Baccarani graduated from the University of Bologna with a degree in electrical engineering in 1967 and a degree in physics in 1969. He joined Bell Laboratories, Murray Hill, New Jersey, in 1969 as a limited-term member of the technical staff, to work on electron-device processing. In 1970 he became a Research Assistant at the University of Bologna, where he investigated the physical properties of MOS structures and transport effects in semiconductor materials and devices. Since

1972 Professor Baccarani has been teaching an annual course on quantum electronics, and in 1980 he was appointed Full Professor of Digital Electronics at the University of Bologna. In 1981 he was a Visiting Scientist at the IBM Thomas J. Watson Research Center, where he investigated the feasibility of a quarter-micrometer MOS process from the standpoint of the physical limitations affecting device performance. In 1983 and 1989 he was again a Visiting Scientist at the Thomas J. Watson Research Center, investigating advanced physical models for device simulation and silicon-based heterojunction bipolar transistors, respectively. Since 1983 Dr. Baccarani has headed a group involved in the numerical analysis of semiconductor devices; he has been a co-leader of four successive EEC-supported projects in the area of CAD for VLSI. His current research interests include numerical-device simulation, synthesis of analog circuits, analog and digital architectures for image processing, and integrated-circuit design.

Antonio Gnudi *Dipartimento di Elettronica, University of Bologna, viale Risorgimento 2, 40136 Bologna, Italy (gnudi%bodem7.cineca.it at iceneca.cineca.it).* Dr. Gnudi graduated *cum laude* with a degree in electrical engineering in 1983 from the University of Bologna, Italy, and received a Ph.D. degree in electrical engineering from the same university in 1988. His dissertation involved the numerical modeling of semiconductor devices. From 1988 to 1990 he was a Visiting Scientist at the Mathematical Sciences Department of the IBM Thomas J. Watson Research Center, where he was engaged in modeling submicrometer devices and associated higher-order moments of the Boltzmann transport equation. Since 1990 Dr. Gnudi has been a member of the Department of Electrical Engineering of the University of Bologna, Italy. His research activity involves low-temperature modeling of short devices, and the use of numerical methods for the solution of the Boltzmann transport equation as it applies to semiconductors.

Johannes M. C. (Hans) Stork *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (STORK at WATSON, stork@watson.ibm.com).* Dr. Stork received an Ingenieur degree in electrical engineering from Delft University of Technology in 1978 and a Ph.D. degree from Stanford University in 1982. His thesis work at Stanford involved physical and technological aspects of bipolar structures for VLSI. During the academic year 1977-1978 he worked at the Philips Research Laboratories, Briarcliff Manor, New York, on the modeling of charge storage in MNOS devices. In 1982 he joined IBM at the Thomas J. Watson Research Center to work on advanced bipolar technology. In 1985 he focused on the technology and design of exploratory bipolar devices, and, in 1987, assumed responsibility for managing an exploratory devices group. In 1990 he managed a bipolar devices group, and is currently responsible for the Bipolar/BiCMOS Technology research program at the Thomas J. Watson Research Center. Dr. Stork has been awarded two IBM Outstanding Technical Achievement Awards and is a Senior Member of the IEEE Electron Devices Society. He has served as a member of the 1988 BCTM program committee, and has been on the VLSI Technology Symposium program committee since 1986. He was publications/publicity chairman for the 1990 and 1991 Technology and Circuit Symposia, and publicity vice-chairman for the 1991 International Electron Devices Meeting.

Jack A. Mandelman *IBM Technology Products, Semiconductor Research and Development Center, Hopewell Junction, New York 12533 (U6641 at FSHVMFK1).* Dr. Mandelman received B.E.E. (*magna cum laude*) and M.E.E. degrees from the City College of New York in 1969 and 1971, respectively. In 1975 he received a Ph.D. (E.E.) degree from the City University of New York, where his doctoral research involved the design, fabrication, modeling, and characterization of distributed microwave diodes. From 1974 to 1975, he was employed by General Instrument Corporation's Microelectronics Division, where he was responsible for characterization and yield enhancement of MOS products. In 1975, he joined IBM in Burlington as a circuit designer in a group focusing on high-performance DRAMs. He has since been extensively involved in the use of advanced device and process modeling for the solution of technology design and reliability problems. He has recently transferred to the Semiconductor Research and Development Center in Hopewell Junction, New York, to work on 256MB DRAM device design. Dr. Mandelman is a member of the Institute of Electrical and Electronics Engineers.

superconductivity. He subsequently worked in the areas of foundations of scattering and band theory, bifurcation theory, numerical stability for stiff equations, and circuit and device analysis. He was an Associate Professor at the American University, Beirut, Lebanon, in 1968, and a visiting member of the Courant Institute, New York, in 1963 and 1976. At the time of his death, Dr. Odeh managed the Differential Equations group in the Mathematical Sciences Department at the Thomas J. Watson Research Center.

Emmanuel F. Crabbé *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (CRABBE at YKTVMV, crabb@watson.ibm.com).* Dr. Crabbé obtained a B.S. degree in electrical engineering from the University of Missouri-Columbia. He received M.S. and Ph.D. degrees in electrical engineering from Stanford University in 1983 and 1988, respectively. The subject of his dissertation was the study of transport in polysilicon emitter contacts for bipolar transistors. He has been with IBM at the Thomas J. Watson Research Center since 1988, and is currently engaged in research on Si-Ge bipolar transistors and MOSFETs. Dr. Crabbé is a member of the American Physical Society and is currently serving on the program committees of the Device Research Conference and the European Solid State Device Research Conference.

Matthew R. Wordeman *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (WORDEMA at FSHVM1).* Dr. Wordeman received a B.S.E.E. degree in 1978 from the University of Cincinnati, an M.S.E.E. degree in 1981 from the Polytechnic Institute of New York, and an Eng.Sc.D. degree from Columbia University in 1986. Since joining IBM at the Thomas J. Watson Research Center in 1978, he has worked on radiation-induced soft errors in miniaturized memories and logic circuits, scaling and nonscaling effects leading to the limits of MOSFET miniaturization, modeling of low-voltage miniaturized MOSFETs, techniques for characterization of MOSFET physical parameters, the operation of MOSFETs at cryogenic temperatures, and the development of submicron CMOS processes. Dr. Wordeman currently manages the DRAM Design group at the Thomas J. Watson Research Center.

Farouk Odeh *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (deceased).* Dr. Odeh received a Ph.D. degree from the University of California in 1961, joining the IBM Research Division to work on mathematical aspects of