RC 20597 (10/29/96)
Engineering and Technology

IBM Research Report

SEMICONDUCTOR DEVICE PHYSICS AND THE MODELING OF SMALL SEMICONDUCTOR DEVICES

S. E. Laux and M. V. Fischetti

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

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

SEMICONDUCTOR DEVICE PHYSICS AND THE MODELING OF SMALL SEMICONDUCTOR DEVICES

S.E. LAUX[a] and M.V. FISCHETTI
IBM Research Division, T.J. Watson Research Center
P.O. Box 218, Yorktown Heights, NY 10598, USA

The symbiotic relationship between modeling the underlying physics in semiconductors and modeling small semiconductor devices is discussed. Both modeling activities depend on large scale computing in an intrinsic way, with progress tied to the continued growth in raw computing power. Modeling the underlying physics in semiconductors involves the numerical evaluation of electronic material properties from detailed physical theories, *e.g.*, bandstructure, scattering rates and conductivity. Device modeling takes the results of these materials calculations and incorporates them into a device context, *i.e.*, computational domains representing a device geometry complete with doping profiles, contacts, and external electrical and/or optical excitations. The mutual benefit between these disciplines is obvious: predictive device modeling benefits from the physics modeling, since improved physical models yield more predictive device models; physics modeling often needs a device context to allow the theory to be directly compared with experimental data derived from device measurements. This chapter reviews our work in both of these areas.

1 Introduction

The effects of the miniaturization of semiconductor devices and circuits can hardly be underestimated. It is a powerful economic force, enabling ever-faster processors connected to denser memory, pervasive wireless connectivity and impressive gains in information storage and exchange. It is almost taken for granted that end-user cost will remain fixed, or even decrease, in spite of yearly improvements in technology. With this push towards miniaturization comes a huge increase in facilities and manufacturing costs, with increasing demands on manufacturability and tolerance control. Numerical modeling plays an important role at every level of device manufacture: from assessing the performance of individual transistors, to circuits and systems, to the operation of the manufacturing facility itself.

The National Technology Roadmap for Semiconductors 1994 revision [1] is a Semiconductor Industry Association report published by SEMATECH, Inc., which documents the anticipated progress in semiconductor miniaturization, and the key problems to be solved in accomplishing this herculean feat. One

[a] currently on sabbatical leave at the Beckman Institute for Advanced Science and Technology, University of Illinois, 405 N. Mathews Ave., Urbana, IL 61801.

topic which permeates the discussion (a so-called "crosscutting technology") is improved modeling and simulation, for all aspects of technology definition and production. The relevant message for this chapter is: the prediction of improved electrical behavior of individual transistors is the foundation upon which a successful new technology is built. In fact, without the promise of improved device performance, the juggernaut of miniaturization would likely be brought to a halt: costs of designing and manufacturing a new technology could not be absorbed if there was no promise of improved product performance. Thus, the predictive modeling of individual semiconductor devices is a pivotal exercise. Such modeling is not without its challenges.

Progress in numerical semiconductor device modeling has taken two parallel tracks: improvements in modeling geometrically complex device structures, and improvements in the underlying physical basis upon which such modeling depends. Both trends are readily gleaned from relevant conference proceedings (*e.g.*, VPAD,[2] NUPAD,[3] and SISDEP[4] conference series, now unified under the name SISPAD;[5] or the IWCE[6,7] conference series). The later track is the focus of this chapter, namely, the task of improving the physical basis of semiconductor device simulation, although the former track cannot be entirely ignored in this discussion. In fact, it is expected that these two tracks will one day be united: the creation of a single computational tool which represents the the best available techniques for incorporating an advanced physical basis in a complicated, three-dimensional device structure. At the present time, however, this unification is hostage to the overwhelming computational demands of each track individually.

The drift-diffusion approximation[8] is the most widely used physical basis incorporated into semiconductor device modeling programs. Many device modeling programs have been developed which treat two- and three-dimensional device geometries at this level of approximation.[9] Silicon devices with feature sizes of 0.1 μm are forecast to be in production by the year 2007,[1] and the validity of the drift-diffusion equations (DDE) at these dimensions is widely questioned.[10] Such doubts as to the validity of the drift-diffusion approximation generally center around the approximation of averaging over many collisions to define meaningful relaxation times and mobilities, and representing collective behavior of charge carriers via an assumed, "near-equilibrium" form of the distribution function. Such simplifications are incorrect as 0.1 μm dimensions are approached: the distance between scattering events can be a significant fraction of the active device region, resulting in "off-equilibrium" distributions of significantly energetic, or "hot," carriers. Thus, better models have needed to revisit the origins of the drift-diffusion approximation and embrace a more rigorous basis.

The drift-diffusion approximation derives from the Boltzmann transport equation (BTE) as applied to transport in semiconductors, and is a semi-classical theory. This means there are aspects of classical and quantum mechanics included in the theory. These aspects are crudely summarized here; see standard texts on solid-state physics for more information.[11–13] Classical mechanics is clearly at the foundation of the BTE: (i) Particles are idealized as "billiard balls" — having well defined position and momentum — whose motion is controlled by Newton's laws. This classical description is based on the many results of the quantum theory of solids[12] (the correspondence principle, the Bloch and acceleration theorems, the "envelope wavefunction" approximation, etc.), which make it possible to describe the wavepackets associated with the electrons as classical particles, at the price of employing an energy-momentum dispersion significantly more complicated than the parabolic law of free particles. (ii) This idealization of a wavepacket as a point-like classical particle is possible only if, in addition, the dimensions of the structure of interest and the length-scale over which the electrostatic potential varies significantly are larger than the wavelength of the "envelope wavefunction." (iii) Scattering processes are "instantaneous," "complete," and "uncorrelated" point-like events, and so ignore the quantum mechanical interference among scattering waves. But quantum mechanics is not forgotten altogether: (i) The energy-momentum dispersion relation is computed quantum mechanically. (ii) The probability per unit time of each scattering process is obtained using Fermi's Golden Rule and the quantum mechanical matrix elements between ingoing and outgoing (envelope) plane-waves. (iii) Quantum size effects (carrier confinement into reduced dimensionality channels, quantum wells, wires or dots) are accounted for by solving the Schrödinger equation in the direction of confinement and using these envelope functions in a reduced-dimensionality BTE describing the remaining two directions. Such solutions to the BTE already tax modern computer resources, and become even more formidable when this transport theory is incorporated into a "realistic" device context. For this reason, virtually all models which solve the complete BTE stop short of three-dimensional device modeling (with the notable exception of the work from Oregon State University,[14,15] and from the University of California at Berkeley[16], where parallel computing is invoked to deal with the size of the calculation).

Beyond the BTE is quantum transport theory.[17,18] While it is tempting to proceed directly to this level of rigor in simulating "small" devices, it is too bold a step. Calculations at this level of rigor[6] often contain ingredients which are repellent to a device technologist: carrier flow (current) may be absent, or only present in the complete absence of scattering; cryogenic temperatures; lack of obvious logical function or control in simulated structures. While it is easy to blame these inadequacies on insufficient computing power (and this is true), it is also true that a far more incomplete picture of the theory necessary for such simulations exists today. In this case, theoretical, as well as computational, progress is necessary before the "state-of-the-art" in modeling quantum transport can be advanced.

This introductory discussion leads to the scope of this chapter. The BTE is solved in a "realistic" device context, using the DAMOCLES™ program.[19–32] This program implements detailed physical models in an effort to include all of the relevant physics for predictive device simulation, and to allow the resulting computer code to act as a "physics benchmark" against which other, more expedient and less rigorous implementations, may be compared. The success of the approach depends heavily on the quality and accuracy of the underlying physical models. Clearly, device modeling depends on the constituent physics. This theme will be considered in the first part of this chapter.

Conversely, detailed physics also depends on device modeling. That is, formulating and calibrating physical models hinges on comparisons between the theory and experiment. And experiments are necessarily performed on semiconductor samples, often themselves complicated semiconductor device structures. Care must be taken to insure the information deconvolved from the raw experimental data is precisely what is desired. Accordingly, establishing accurate physical models is often most readily accomplished in a device modeling context: in this way, external experimental conditions are automatically included into the calibration process. This notion of how physics calibration depends on device modeling is the subject of the second part of this chapter.

Throughout this chapter, the discussion centers on silicon device technology and especially, the Si metal-oxide-semiconductor field-effect transistor (MOSFET). This is in keeping with our interest in and relationship to digital computer technology.

2 Device Modeling Depends on Physics Modeling

The assertion that device modeling depends on the underlying physics is hardly shocking. The intention here is to quantify this statement, using two examples: the effect of the bandstructure model and the calibration of the bulk transport model.

2.1 Effect of the bandstructure model

Having settled on the need to solve the BTE (going beyond the too simple drift-diffusion description and stopping short of the unwieldy full quantum transport approach), the basic ingredients of the transport model need to be ascertained. Namely, what form of the band structure should be used? Circa 1988 this was an important question: parabolic and non-parabolic band structure models for bulk Si were well established [33] and full band models were under active investigation. [34,20] Each band structure model came complete with its own recommended set of scattering rates which were selected to reproduce electron velocity-field data in bulk Si. As can be seen in Fig. 1, the scattering rates

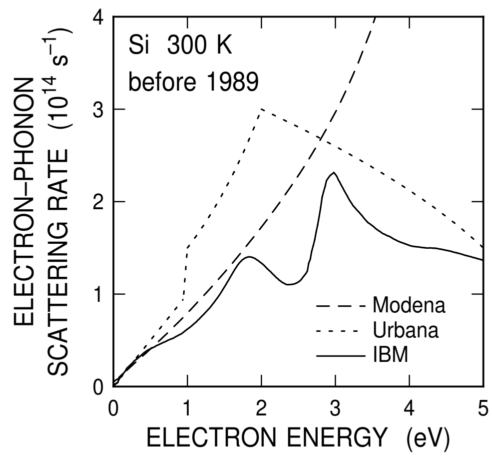

Figure 1: Electron-phonon scattering rates prior to 1989 were in poor agreement at high energies. The curve marked "Modena" is obtained from a parabolic approximation to the band structure, while the remaining two curves were associated with full band models. Figure adapted from Ref. 31.

used by the parabolic band and two full band models were quite different, a reflection of the very different density-of-states (DOS) of the three models. Yet, all models tend to the same behavior at low energy, as required to fit the velocity-field data. This is somewhat disconcerting: device calculations seek to understand hot carriers and their influence on device behavior, yet these early models can be expected to differ substantially in their quantitative predictions at high energies. This very naturally lead to questions surrounding the calibration of band structures and scattering rates, which will be further discussed in Sec. 2.2.

Deferring the question of calibration to Sec. 2.2, first consider whether the issue of calibration is an important one, *i.e.*, how sensitive are device behaviors to the popular transport modeling parameters of the time? Fig. 2 compares the average kinetic energy, drift velocity and electron distribution function near

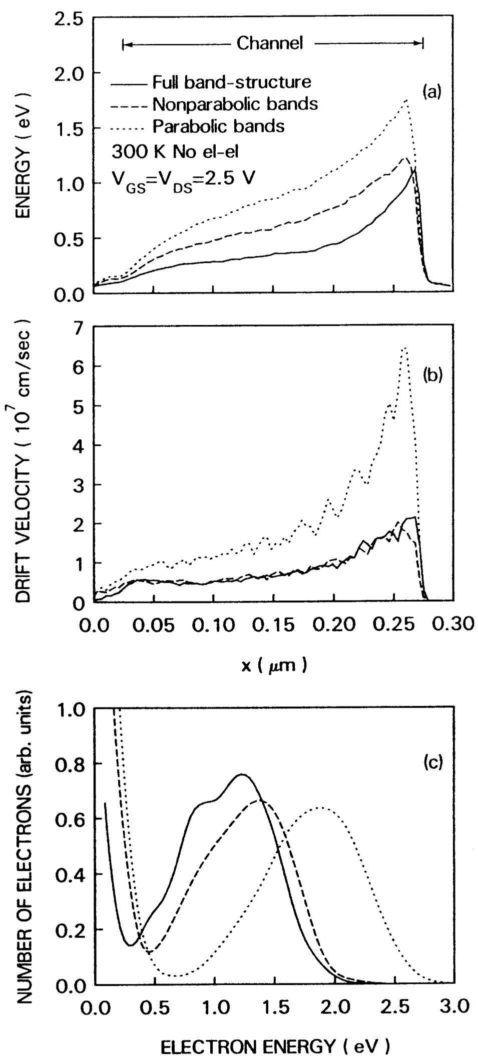

Figure 2: A comparison between results obtained for the same device and bias, but with different transport models. Each transport model has the band structure type as indicated, and an associated set of scattering rates fitted to bulk velocity-field data in Si at 300 K. In (a), the kinetic energy versus position from source (left) to drain (right) in the 0.25 μm nMOSFET depends sensitively on the model employed. In (b), a similar plot for drift velocity indicates the parabolic band model predicts velocities substantially higher than the other models. In (c), the distribution of electron energies in the vicinity of the drain of the nMOSFET is shown: the band structure and its associated density-of-states is reflected in the shape of the distribution function. Figure adapted from Ref. 20.

the drain of a 0.25 μm n-channel Si MOSFET for three transport models.[20] The variations are substantial, and are due to the different underlying physics.

Three models are generally used to describe the energy-wavevector relationship (the "band structure") of the semiconductor: A "parabolic model," which retains only the leading (quadratic) term, E_p, in the expansion of the "exact" energy E in powers of the wavevector $\mathbf{k}$:

$$E_p(\mathbf{k}) \; = \; \frac{\hbar^2 \mathbf{k}^2}{2m^*}, \tag{1}$$

where $\hbar$ is the Planck constant divided by 2π. Eq. (1) expresses the energy-momentum relation of a free electron, having replaced the electron mass, m, with an "effective mass," m^*. This simple approximation is satisfactory at low energies. Therefore, its use is justified in the context of "cold" quasi-equilibrium electrons, as in drift-diffusion approaches. But as electrons become "warm," higher-order terms in the expansion of E in powers of $\mathbf{k}$ become significant. The "non-parabolic" dispersion, $E_{np}(\mathbf{k})$, retains the next leading term (of order k^4):

$$E_{np}(\mathbf{k}) \; = \; E_p(\mathbf{k})[1 - \alpha E_p(\mathbf{k})], \tag{2}$$

where α is a "nonparabolicity parameter," usually positive, which depresses the value of E_{np} with respect to E_p at a given $\mathbf{k}$. The value of α is inversely proportional to the energy gap of the semiconductor. Finally, "full band" models refer to those band-structure models obtained using either local[35] or nonlocal[36] empirical pseudopotentials. These calculations are targeted at obtaining the most accurate excitation spectrum of the semiconductor at the price of using empirical form factors for the pseudopotentials. Space limitations prevent us from entering into a detailed description of how these full band models are implemented numerically. Once this is completed, large look-up tables are generated in order to obtain by interpolation the energy, $E_\nu(\mathbf{k})$, and group velocity, $\nabla_\mathbf{k} E_\nu(\mathbf{k})/\hbar$, of an electron at a point $\mathbf{k}$ in the first Brillouin Zone of the crystal and in the band ν. The selection of the final state after a collision, requiring a search over complicated equi-energy surfaces, remains the computationally challenging step of all programs employing this sophisticated band-structure model.

Returning to the results of Fig. 2, the parabolic band model predicts the "hottest" electron behavior: for drain-source bias $V_{DS} = 2.5$ V, average energies exceed 1.5 eV and the velocity exceeds 6×10^7 cm/sec at the drain end of the channel. The non-parabolic model predicts somewhat lower average energies, and markedly lower velocities. The full band model gives lower average energies still, but agrees closely with the drift velocity given by the non-parabolic model. Overall, this is expected: inclusion of the non-parabolic correction brings the results into closer agreement with the full band model. The two velocities match so well because velocity-field data were used to calibrate the models in the first place. From the device perspective, the differences in the distribution functions shown in Fig. 2(c) are most dramatic. For example, the models predict vastly different numbers of electrons above 1.5 eV. This difference will alter the number of electrons available to impact ionize. Also, the number of very low energy electrons ($\leq$ 0.3 eV) is quite different, even though the same spatial region near the drain was sampled for all three models. These low energy electrons are those electrons which reside in the drain and have not traversed the channel. The full band model predicts a lower diffusion coefficient, which causes the drain electrons to be more tightly confined to the drain itself. Thus, there are relatively fewer low energy electrons in a sample space which straddles the drain-channel junction. This lower diffusion constant is also seen in Fig. 2(a), as a steeper drop in average energy as the drain is entered for the full band model.

Modeling shorter channel length devices, or decreasing the ambient temperature, exacerbates the off-equilibrium nature of the transport in the MOSFET channel. This occurs because either will result in fewer scattering events in the device channel, allowing electrons to attain velocities far in excess of the saturated velocity in Si ($\sim 1 \times 10^7$ cm/sec at 300 K and $\sim 1.2 \times 10^7$ cm/sec at 77 K). This is nicely indicated in Fig. 3, where the same three plots shown in Fig. 2 are repeated here, for a 0.06 μm channel length nMOSFET simulated at 300 and 77 K. The possibility for "velocity overshoot" is clearly indicated in Fig. 3(b), where at 77 K electrons exceed the saturated velocity over virtually the entire channel. Also note how energies are higher at 77 K, indicative that hot electron effects increase as temperature decreases, due to an increase in the mean free path length which allows electrons to gain more energy between scatterings.

In spite of the early question concerning calibration of bands and rates, a result which bolstered our confidence in the DAMOCLES™ model is shown in Fig. 4. When compared to experimental transconductance data (an important figure of merit, strictly related to the switching speed of devices in logic circuits) versus channel length at 300 and 77 K, the modeling performed admirably [19,20] when compared to the data of Sai-Halasz, *et al.* [37] The experimental evidence for velocity overshoot, and the good agreement of the model to the data, was an important first result. It implied the model apparently possessed elements of truth, in spite of the difficulty and uncertainty in calibrating the model at that time, as discussed in Sec. 2.2.

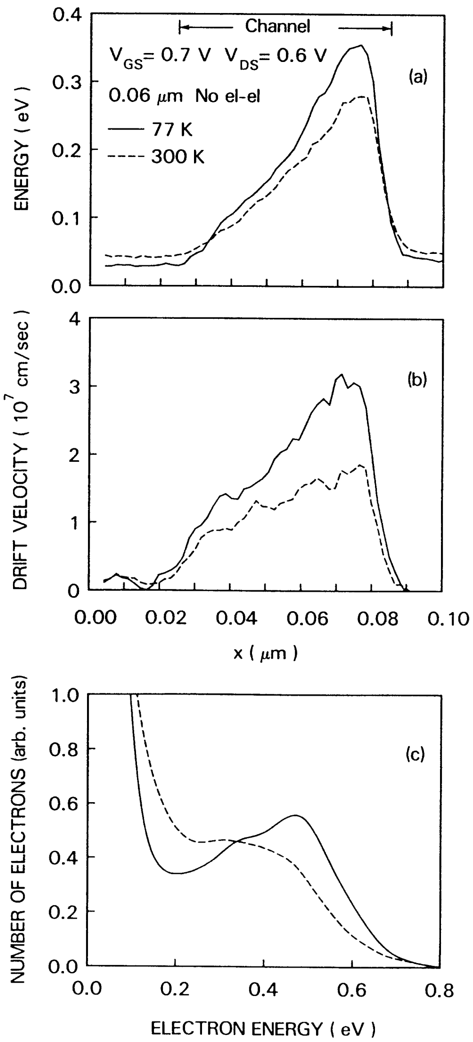

Figure 3: Temperature dependence of the (a) electron energy, (b) drift velocity and (c) distribution function for a 0.06 μm nMOSFET, for the bias as indicated. Off equilibrium effects and electron heating are greater as the temperature is decreased, because the scattering rate decreases as temperature decreases. The resultant device behavior thereby approaches "ballistic" (scattering-free) transport more closely. Figure adapted from Ref. 20.

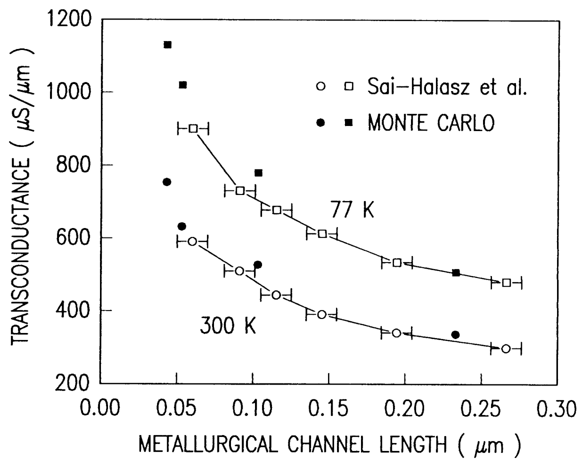

Figure 4: Experimental and simulated transconductance versus metallurgical channel length for nMOSFETs at 300 and 77 K. The large transconductance values at the smaller channel lengths, especially at 77 K, can only be explained if the channel velocity exceeds the saturated velocity in Si. Figure adapted from Ref. 20.

2.2 Calibration of bulk energy bands and scattering rates

A model of bulk transport in a semiconductor should correctly reproduce bulk transport properties, notably, carrier velocity versus electric field and impact ionization coefficient versus electric field. The two main ingredients of such a model are the band structure and the scattering rates. As the paper of Abramo, *et al.*[38] clearly demonstrated, a wide span of model parameters has appeared in the literature to describe Si bulk transport. Such a wide variance must be addressed prior to maintaining predictive device simulation is possible.

Fortunately, this same paper had the effect of beginning to narrow down the number of models considered reasonable representations of the physics. From the point of view of band structure, empirical (but accurate) pseudopotentials are now considered necessary: the density-of-states is accurately represented by these bands, which in turn controls the gross features of the carrier-phonon scattering rates. While uncertainty still remains in the exact form of the Si band structure at high energy (>1.0 eV), empirical pseudopotentials seem adequate for the task at hand. Clearly, additional experimental verification of band splittings and fine structure are welcomed, especially away from valley minima.

Pinning down phonon and impact ionization scattering rates has proven more difficult, even after the form of the bands has been fixed. Actually, even a cursory glance at the literature on electron transport in Si will reveal the surprising fact that, after more than four decades of study, a huge uncertainty still affects our description of even the gross features of scattering probabilities at large (above 0.2 eV or so) electron energies.[28] Only recently, as shall be seen, a "consensus" has been reached. The reasons for this embarrassing delay in understanding quantitatively the strength of the scattering processes in Si is ultimately caused by the daunting difficulty of the task. Consider first the electron-phonon interaction. Even embracing already restrictive simplifications (harmonic coupling, Born-Oppenheimer and rigid-pseudoion approximations, Fermi's Golden Rule), the rate $1/\tau_{\eta}^{\pm}(\mathbf{k},\nu)$, at which an electron in band ν, with quasi-momentum $\hbar\mathbf{k}$ will emit (upper sign) or absorb a phonon of wavevector $\mathbf{q}$ ($= \mathbf{k}-\mathbf{k}'$), of mode (acoustic/optical, longitudinal/transverse) η and frequency $\omega_{\mathbf{q}\eta}$ will be:

$$\frac{1}{\tau_{\eta}^{\pm}(\mathbf{k},\nu)} = \frac{2\pi}{\hbar} \sum_{\nu',\mathbf{k}'} |M_{ep}^{\pm}(\mathbf{k}',\nu';\mathbf{k},\nu)|^2 \ \delta[E_{\nu}(\mathbf{k}) - E_{\nu'}(\mathbf{k}') \pm \hbar\omega_{\mathbf{q}\eta}]. \quad (3)$$

The matrix element $M_{ep}^{\pm}(\mathbf{k}',\nu';\mathbf{k},\nu)$ can be expressed in terms of the mass of the primitive cell, M, of the number of cells, N, and of the Bose occupation number of phonons, $N_{\mathbf{q}\eta}$:

$$M_{ep}^{\pm}(\mathbf{k}',\nu';\mathbf{k},\nu) = \left(\frac{\hbar}{2NM\omega_{\mathbf{q}\eta}}\right)^{1/2} \Delta_{\eta}(\mathbf{k}'\nu';\mathbf{k}\nu) \ (N_{\mathbf{q}\eta} \mp 1/2 + 1/2)^{1/2}, \quad (4)$$

where the quantity Δ_{η}, known as the "deformation potential," depends on the ionic potential and on the phonon polarization vectors. The former, even if assumed to be rigidly shifted and "pseudized," is very poorly known, while the latter quantities depend on lattice dynamics models fitted to phonon spectra (the eigenvalues), but which provide no guarantee about the quality of the eigenvectors (the all-important polarization vectors). At low electron energies, the Bardeen and Shockley[39] connection between the macroscopic shifts of the band or valley extrema (related to the "proper" deformation potential, and easily measurable) and Δ_{η} comes to rescue. But at higher energies very little was known.

A similar situation had to be confronted when tackling the process of pair-creation across the gap, *i.e.*, the process by which an electron with kinetic energy exceeding the energy-gap excites a valence-electron across the gap, recoiling to a lower-energy state and generating an electron-hole pair. This process, the inverse of the well-known Auger process, is of foremost importance in devices, since it controls the (avalanche) breakdown strength of the material.

Within the Born approximation, the rate for this process, $1/\tau_{ii}(\mathbf{k}\nu)$, initiated by an electron of quasi-momentum $\hbar\mathbf{k}$ in band ν can be calculated as:

$$\frac{1}{\tau_{ii}(\mathbf{k}\nu)} = \frac{2\pi}{\hbar} \sum_{\mathbf{GG}'} \sum_{\nu_v \nu'_c \nu''_c} \int \frac{d\mathbf{k}'}{(2\pi)^3} \int \frac{d\mathbf{p}'}{(2\pi)^3} \, |M_{\mathbf{GG}'}(\mathbf{k}\nu, \mathbf{p}\nu_v; \mathbf{k}'\nu'_c, \mathbf{p}'\nu''_c)|^2$$
$$\times\delta[E_\nu(\mathbf{k}) + E_{\nu_v}(\mathbf{p}) - E_{\nu'_c}(\mathbf{k}') - E_{\nu''_c}(\mathbf{p}')], \qquad (5)$$

where M is the (anti-)symmetrized screened Coulomb matrix element obtained by adding the scattering probabilities in the singlet and triplet states. Here one must confront a hard computational problem caused by the large dimension of the integrals in Eq. (5) and by the necessity of employing full band models for the dispersion, as dictated by the large energies involved. In addition, the Coulomb potential is obviously screened, with dynamic and wave-length-dependent corrections clearly non-negligible, in light of the large energy and momentum transfers involved in the process. Again, a daunting physical and numerical task. Considering the difficulty posed by these semi-rigorous approaches, it should not come as a surprise that empirical methods have been used for many years: By employing empirical extrapolations of low-energy (such as the Shockley-Bardeen deformation potentials) or oversimplified (such as the Keldysh formulation of the pair-production rate, valid in the limit of parabolic and direct-gap band-structures) concepts beyond their range of validity, attempts were made to extract the energy-dependence of the scattering process from fits to experimental data such as drift-velocity and ionization coefficient versus field characteristics. This is the framework referred to as the calibration problem. Unfortunately, several unknowns (all electron-phonon deformation potentials and the ionization rate) must be determined via a comparison with scarce experimental information. In addition, some of the unknowns, such as the ionization rate, appear to be only loosely constrained by the available experimental data, rendering the situation quite discouraging.

Fig. 5 shows the last predicament quite clearly. The same band structure is used for two calculations of the electron distribution function and the ionization coefficient in bulk Si. A constant field of 300 kV/cm in the (100) crystal direction is specified. Most importantly, the same electron-phonon scattering rates are used in the two calculations: the only change is that two different impact ionization rates are used, as shown in Fig. 5(a). The results are quite disconcerting: in spite of the large change in the number of energetic electrons above 3 eV, the corresponding ionization coefficients both agree well with the experimental data, as seen in Fig. 5(b). The reason is simple enough: a larger high energy ionization rate "cools" the distribution profoundly at high energies, as electrons are more rapidly undergoing the energy loss of impact ionization.

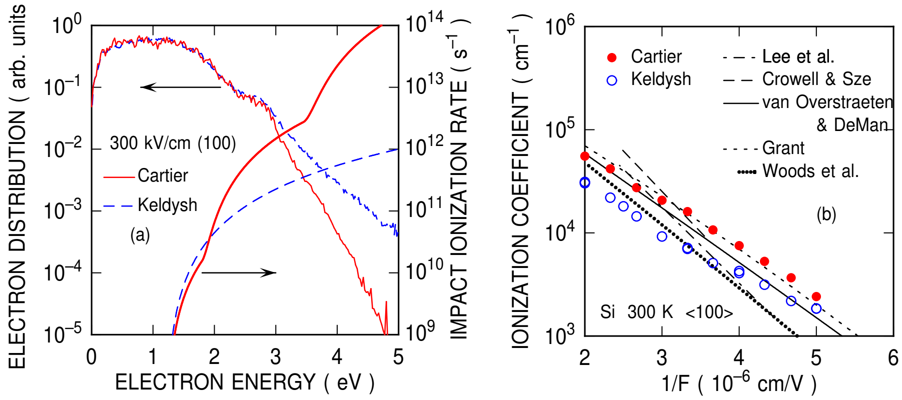

Figure 5: For a fixed set of electron-phonon scattering rates in bulk Si, two different impact ionization rates produce significantly different high energy populations, as increasing the impact ionization rate has a "cooling" effect on the high energy tail of the distribution, as seen in (a). However, the ionization coefficient, which roughly derives from the product of the number of energetic electrons present and the rate of impact ionization, is almost unchanged, as seen in (b). The ease of fitting the ionization coefficient is unfortunate and does not aid in establishing the impact ionization rate uniquely.

However, there are then fewer electrons to undergo impact ionization in the first place, nearly canceling out the expected increase in the ionization coefficient; the ionization coefficient only increases a relatively small amount. Note the agreement with bulk velocity-field data for both calculations is essentially identical; these high-energy calibrations do not effect the lower energy behavior of the models. The key piece of high-energy data — the ionization coefficient — proves to be too easy to fit, given its sublinear dependence on the impact ionization rate! How then can the impact ionization rate be ascertained? Note this quandary was clearly identified even in early full band Monte Carlo research.[34]

Recently, this difficultly in calibration has seen much progress, to the point where the bulk transport model is certainly ready for device applications. This progress has come from multiple direction: (i) *ab initio* calculations of electron-phonon scattering rates;[40,41] (ii) *ab initio* computations of impact ionization rates;[32] and (iii) new experimental data for impact ionization rates deduced via Monte Carlo transport calculations.[42] The consensus for the electron-phonon and impact ionization scattering rates in Si, as a result of these three complementary investigations, has been quite satisfactory. Fig. 6 shown compellingly the level of agreement present in the recent literature (post 1992) for

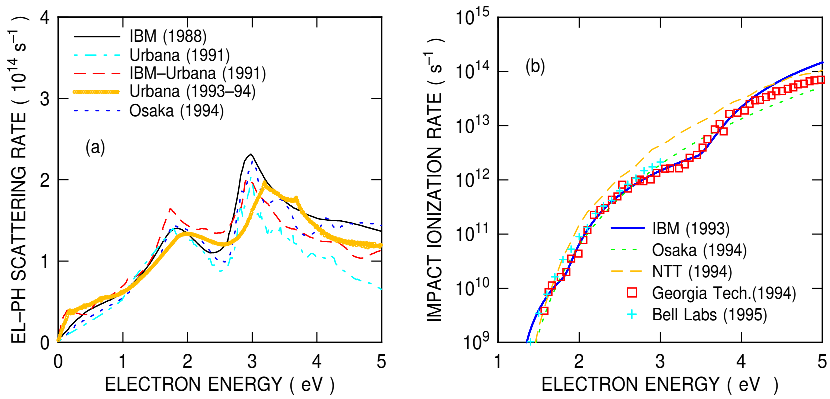

Figure 6: A new consensus has evolved in the recent literature (post 1992) for the (a) electron-phonon and (b) impact ionization rates for electrons in Si. Figures adapted from Ref. 31.

the electron-phonon and impact ionization scattering rates in Si. In sum, the early criticisms that too many parameters known with too little certainty were at the root of every full band Monte Carlo bulk transport model no longer apply, at least for electrons in silicon. As one strays from this material, or even begins to consider hole transport in any material, the measure of confidence is lower, but should be forthcoming as researchers continue to combine modeling and experiments towards this goal.

3 Physics Modeling Depends on Device Modeling

This section discusses how the calibration of relevant device physics is facilitated by the preexisting device modeling context. In this case, experimental results (often aimed at clarifying the underlying physics) were obtained from device structures of nontrival complexity, and the most straightforward way to understand the experiments was to simulate the entire device. By investigating computationally various combinations and calibrations of the underlying physical models, a greater degree of confidence in the underlying models results. Three examples are discussed: (i) energetic electrons at the Si-SiO_2 interface; (ii) low energy electrons and quantization effects at the Si-SiO_2 interface; and (iii) the importance of dynamic screening in electron-electron scattering.

3.1 Energetic electrons at the Si-SiO_2 interface

The behavior of energetic electrons at the Si-SiO_2 interface is extremely complicated.[28] As energetic electrons approach the $\sim$3.1 eV barrier from Si into SiO_2, the possibility exists for the electron to be injected from Si into the SiO_2. This injection process is influenced by many details, such as conservation of energy and momentum as the interface is crossed, scattering in the SiO_2 itself, the effects of the image force, and the role of tunneling in facilitating injection. These physics are needed to understand, and hopefully predict, the flow of gate current in nMOSFETs from the incoming flux of hot electrons entering the SiO_2 layer. From the incoming flux, one hopes to be able to ascertain some information about the rate of device degradation due to the trapping of electrons in the insulator. Such device degradation can ultimately cause an electrical circuit or logic gate to cease functioning, and is hence of great technological importance.

The seminal experiments on injection of energetic electrons from Si into SiO_2 were done by Ning, *et al.*[43] The device structure employed consisted of a (long-channel) nMOSFET. The source and drain contacts were grounded; the substrate contact was negatively biased and the gate contact was positively biased to insure the formation of an inversion layer of electrons and establish the channel potential as equal to the drain/source potential. Electrons were photogenerated and accelerated towards the Si-SiO_2 interface in the field created by the substrate bias; electrons then enter the SiO_2 if they gain sufficient energy from the field, or via tunneling. The ratio of the gate current to the drain+source current gives the probability of injection at a given bias. This experiment gives a sensitive monitor of the accuracy of the modeling of energetic electrons at the Si-SiO_2 interface. By simulating a full nMOSFET device geometry, injection probability can be calculated precisely as it is measured in the laboratory. This automatically takes into account the interdependencies of the individual physical models included in the calculation, and greatly facilitates ascertaining the physical models best able to reproduce the experimental data.

Fig. 7 illustrates schematically the band-diagram associated to the experimental situation and indicates some of the troublesome issues with which physicists must deal. As explained in Ref. 28, the amorphous nature of SiO_2 and experiments on tunneling across SiO_2 from Si-substrates of different crystallographic orientation[44] have hinted at the possibility that the component of the electron momentum lying on the plane of the interface (parallel momentum) may not be conserved. Similarly, experiments on photon-assisted tunneling have been interpreted as being consistent with the absence of the

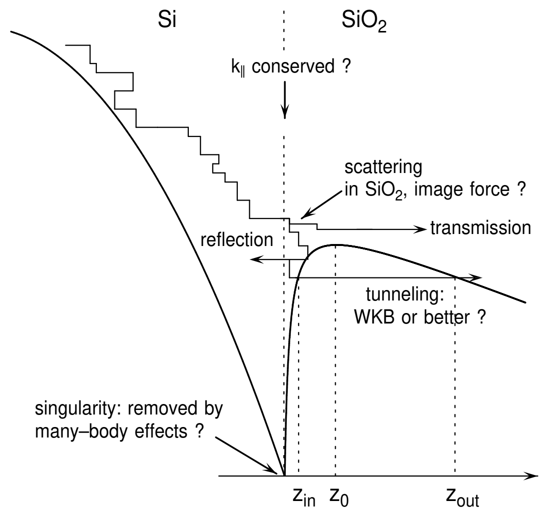

Figure 7: Schematic band diagram of the Si-SiO_2 interface illustrating some of the many physical issues which complicate the understanding of the electron-injection process. Figure adapted from Ref. 28.

classical image-force at the Si-SiO_2 interface.[45] These issues have been debated at length in the literature. While the latter has been clarified by work performed within the context of scanning tunneling microscopy[46], the former remains controversial to this day.

As an example of the procedure of establishing the physical models best able to reproduce the experimental data, consider Fig. 8. This figure shows the measured and simulated injection probability for device number 15-12-8 of Ref. 43 as the gate bias, and hence the oxide field, changes. In Fig. 8(a), two computations are performed, differing in the manner in which momentum conservation is handled as an electron moves from Si into SiO_2. In one case, crystal momentum parallel to the interface is conserved; in the other case, the total energy on the Si side (minus the energy barrier) is directed into the perpendicular momentum in the SiO_2 (labeled total energy conversation), and a WKB approximation is used to evaluate the tunneling probability for electrons below the barrier height. The former physical model compares very favorably with the data, while the later shows substantial disagreement. This gives some confidence relative to the selection of the proper physical model related to energy and momentum conservation at the Si-SiO_2 interface.

Fig. 8(b) is an analogous comparison, this time considering the individ-

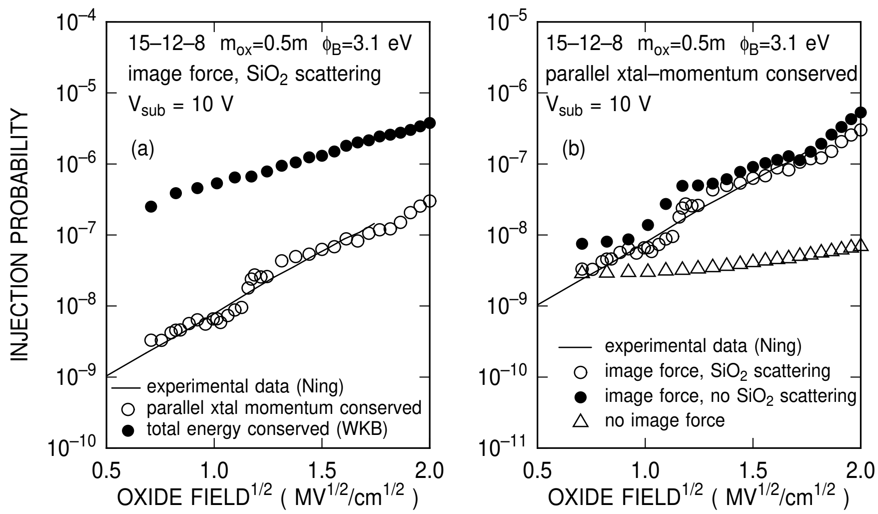

Figure 8: Measured and simulated injection probability versus oxide field for the device 15-12-8 of Ref. 43. (a) Image force lowering and scattering in the SiO_2 are included for both calculations shown. Conservation of parallel crystal momentum at the interface is in far better agreement with the data then assuming total energy conservation at the interface. (b) Various approaches related to the modeling of the image force are compared. Neglecting the image force gives the wrong field dependence of the injection probability. Upon including the image force, scattering in the SiO_2 improves the agreement with the data slightly, compared to neglecting such scattering. Figures adapted from Ref. 28.

ual effect of the barrier lowering due to the image force and the possibility of scattering in the SiO_2 in the interval between the interface and the potential maximum in the SiO_2. Again, comparison with the data demonstrates the necessity of including the image force; the oxide field dependence of the injection probability without it is in poor agreement with the data. Having included barrier lowering due to the image force, further inclusion of scattering in the SiO_2 improves even more the agreement between the simulation and the data.

The ability to directly simulate the Ning device structure and compare the results directly with measured data lends confidence to the physical models employed; the simulation is only possible through the ability to model a complete nMOSFET device. Thus, selection among physical models was facilitated by the device modeling capabilities of the DAMOCLES™ program.

3.2 Low energy electrons and quantization at the Si-SiO$_2$ interface

Having established a bulk transport model for electrons and holes in Si in which there is some confidence[32] — and for electrons, some consensus (Fig. 6) — there is still a major issue related to modeling MOSFETs, the majority of which are "surface channel" devices. This means the channel current flows between source and drain along the silicon surface, or specifically in Si MOSFETs, the Si-SiO$_2$ interface. This introduces additional scattering mechanisms, *e.g.*, surface acoustic and optical modes, Coulomb scattering, and surface roughness scattering.[27] Two approaches have been generally pursued in an effort to establish a predictive model of the interface physics: (i) "partially diffusive" scattering models from classical physics; and (ii) quantization of the surface region combined with simultaneous inclusion of quantum (low energy) and bulk (high energy) transport.

The first approach is based on classical mechanics and it is purely empirical in nature: Electrons hitting the Si-SiO$_2$ interface can be either specularly reflected or diffusively and elastically scattered. The merit of this approach is its simplicity, but its drawback is amply illustrated in Fig. 9: The figure shows the output characteristics for an nMOSFET with a 0.1 μm metallurgical channel length and doping similar to the devices of Ref. 37. Results from standard drift-diffusion simulations are compared with Monte Carlo simulations using varying probabilities, r, of specular scattering (so that $1 - r$ is the probability of diffusive scattering). The major results of a complex simulation appear to be totally at the mercy of a simple parameter with little or no physical basis. Yet, this approach remains widely used. Note, in passing, that the value $r = 0.75$ provides the best agreement between Monte Carlo and calibrated drift-diffusion results at low fields (*i.e.*, low drain-source voltage), while the Monte Carlo results correctly predict larger currents at larger voltages, as a result of the off-equilibrium electron transport (velocity overshoot) in these small devices.

The persistence of this simple approach in the simulation community is justified only by the rather complicated nature of the "correct" approach. For lower energy electrons, the band-bending at the Si-SiO$_2$ interface results in a very narrow potential well which quantizes these electrons. Thus, detailed understanding of how the electrons move from source to drain in the inversion layer requires a far more sophisticated treatment than the "partially diffusive" surface scattering model discussed above. Specifically, quantized electron charge needs to be represented as wavefunctions populating energy subbands. These subbands arise in response to no longer representing the conduction band as a continuum of bulk states, but rather, by combining discrete energy

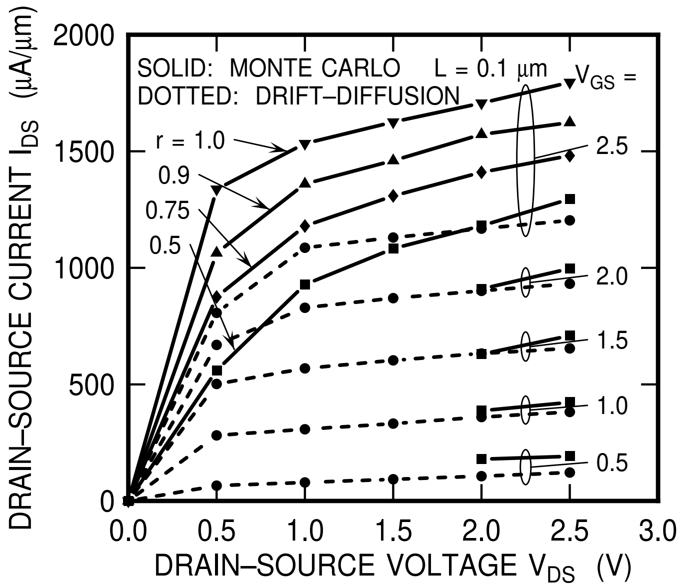

Figure 9: Output characteristics of a 0.1 μm nMOSFET, simulated with drift-diffusion (dotted) and full band Monte Carlo (solid), at 300 K. The Monte Carlo results are parameterized by the probability of specular scattering r at the Si-SiO$_2$ interface; diffusive scattering occurs with probability $1 - r$. All Monte Carlo results for $V_{GS} \leq 2.0$ V pertain to $r = 0.5$. As the probability of specular scattering increases, the drain-source current increases. Monte Carlo and DDE should match as $V_{DS} \to 0$ V; using this criterion, $r = 0.75$ seems most appropriate.

states representing the confinement perpendicular to the Si-SiO$_2$ interface with continuum states in the remaining two space dimensions parallel to the planar Si-SiO$_2$ interface. Bulk transport is labeled "3D" transport, as the continuum approximation is used in all three space dimensions; for the situation at hand, transport with confinement perpendicular to the interface is termed "2D" transport (only two dimensions of bulk transport). 2D/3D transport poses formidable problems within the context of device simulation;[27] however, simplifications can be made in the case of long devices. In this case the potential along the channel varies slowly enough to ignore quantization effects along this direction, thus reducing the solution of the 2D or 3D Schrödinger equation to a 1D problem. Yet, even under these simplifying assumptions, the problem remains a hard numerical task: (i) quantizing in the direction perpendicular to the Si-SiO$_2$ interface dictates the Schrödinger equation must be solved along the perpendicular direction at many positions along the channel, in order to ascertain quantized behavior along the entire channel length (the subbands will vary along the channel subject to the source-drain potential, as well as the gate field). This admixture of quantized and bulk representations is a simplification taken to avoid quantizing in both perpendicular *and parallel* channel directions, which remains out of reach for present-day computation. (ii) Having accepted the idea of quantizing perpendicular to the interface only, it is not appropriate to quantize electrons in the source and drain of the device. Quantization here is not necessary; bulk transport is an excellent approximation in this part of the device. Unfortunately, the 3D transport in the source/drain must be married to the 2D transport in the channel. How to connect these two conduction paths is not obvious, and the origin of more approximation. (iii) Finally, it is not practical even to quantize all electrons in the channel region. As electrons move from source to drain, they gain significant energy from the source-drain field. As they become more energetic, more and more subbands are necessary to fully model the available states at higher energies. Thus, many subbands are necessary in such modeling. Incorporation of so many subbands again outstrips the capabilities of present-day computation, so another approximation is adopted: in the vicinity of the inversion layer, electrons may be quantized (if low in energy) or in bulk states (if sufficiently high in energy). Placing energetic carriers in bulk states is reasonable, as higher kinetic energies imply shorter wavelengths, which relax the need to quantize the electron. However, the detailed interactions between 2D and 3D electron populations must be carefully constructed, in order to permit the physically-mandated intra- and inter-population scatterings. (iv) Finally, physical models for surface scattering for both the 2D and 3D populations needs to be considered, and calibrated. These four areas of approximation hint at the physical complexity of the models necessary to even begin to understand inversion layer transport in short-channel nMOSFETs.

The importance of the device modeling context becomes clear when considering the device structure that was used to exercise the 2D/3D transport model described above. The goal of ascertaining the character of the physical models is best realized by creating an inversion layer where both perpendicular and parallel electric field strengths are as constant as possible. A standard nMOSFET device will not suffice: the perpendicular field will remain constant only if there is no source-to-drain potential variation, *i.e.*, only if the parallel field is zero. Both experimentally, and in our modeling, this difficulty is circumvented by introducing an MOS gate which is *not* an equipotential. That is, the gate potential varies linearly from source to drain in order to maintain a constant perpendicular field along the entire channel length. The results below are obtained for a 1 μm nMOSFET with a non-equipotential gate. The gate is approximated by twenty small gates yielding a stairstep approximation to the desired linear gate potential variation.

Even after the implementation and careful calibration of such an inversion layer transport model for electrons in Si, comparisons between theory and experiment of device-level behavior was, at first, somewhat disappointing.[27] Recent results are finally satisfactory, after many years of puzzlement. Figs. 10 and 11 summarize these results. Fig. 10 shows the effective electron

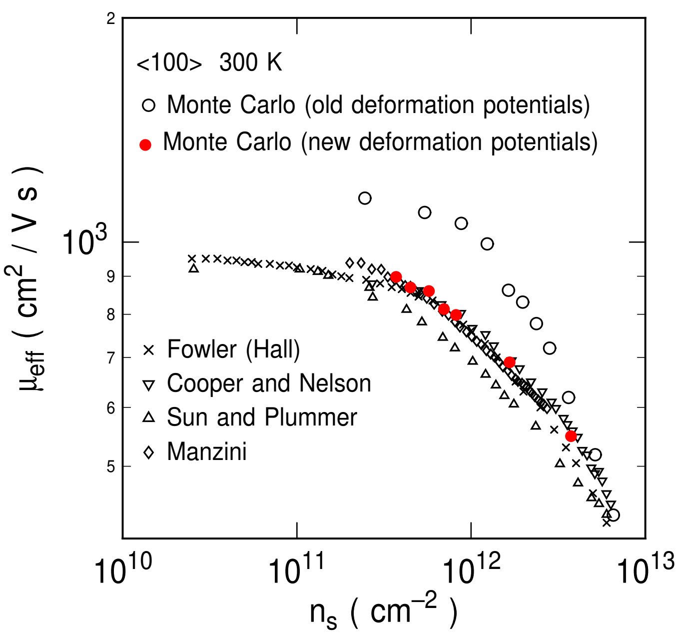

Figure 10: Effective inversion layer mobility versus electron sheet density as calculated with DAMOCLES™ compared to experimental data from multiple sources. The phonon-limited component of the calculated mobility is too large, compared to experiment, when using older deformation potentials from the literature. Recent calculation of these quantities and a revival of older intervalley scattering parameters seem to yield a satisfactory agreement.

mobility in the channel as a function of the electron density in the channel, a quantity obviously related to the field perpendicular to the Si-SiO_2 interface. The experimental data plotted in Fig. 10, as cited in Ref. 27, are obtained in lightly-doped channels, so that impurity scattering has a negligible effect. Also shown are the early results from the DAMOCLES™ program (open circles), as well as some more recent results (solid symbols). The calculated effective mobility was in the past too large compared to the experimental data, especially at low electron densities. The shape of the calculated mobility curve derives from three limiting physical effects: (i) at low electron density, impurity scattering sets the mobility in the inversion layer, as the local electron density decreases to the point where the impurities are no longer screened; (ii) at very high electron concentrations, the correspondingly large perpendicular (to the Si-SiO_2 interface) electric field "crushes" the carriers against the interface and surface-roughness-scattering determines the mobility; and (iii) at intermediate densities, the mobility is essentially determined by phonon scattering. Upon close examination of the open theoretical symbols in Fig. 10, the two limiting behaviors (i) and (ii) are respectably reproduced and in agreement with the data. This implies phonon processes in the inversion layer, seemingly so well quantified in bulk Si, are poorly modeled in the inversion layer. Recent calculations [30] of the scattering parameters (the deformation potentials) for the electron/acoustic-phonon process, benefiting from recent experimental results on the band-structure and electron mobility in strained Si and SiGe alloys, has allowed us to improve dramatically the situation, as shown by the solid circles.

Fig. 11 shows the electron drift velocity in the inversion layer versus par-

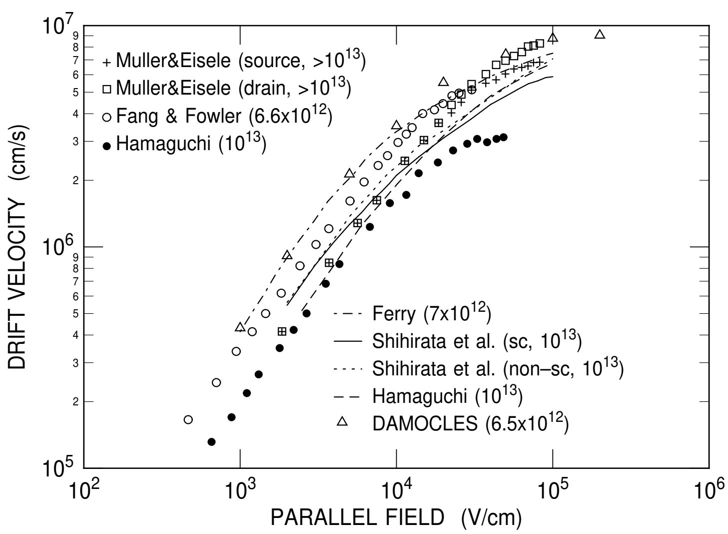

Figure 11: The electron drift velocity in the inversion layer versus electric field parallel to the Si-SiO_2 interface, as calculated here (triangle) and elsewhere (lines), and according to experiments (other symbols). The numbers in parenthesis give the inversion layer density in cm^{-2}. DAMOCLES^{TM} results seem to predict a saturated velocity that is too large. Figure adapted from Ref. 27.

allel field (drain-source field), for a fixed inversion layer density (and hence fixed perpendicular field). These data are particularly relevant if quantitative nMOSFET modeling including the effects of the inversion layer are to be successful. Many sets of experimental data are compared to the results obtained with the DAMOCLES™ program. The computation reveals good agreement with the data for relatively low parallel field values, an expected result since the inversion layer mobility is in good agreement with the data at the effective perpendicular field $F_{eff} \sim 5 \times 10^5$ V/cm used. However, as the parallel field increases and velocity saturation in the inversion layer is anticipated, the Monte Carlo model predicts velocities higher than the data, especially considering the authors' bias towards the data with the lower saturation velocities. The model seems to approach the "bulk" saturation velocity, caused undoubtably by the model's characteristic of not quantizing sufficiently energetic electrons. While these computations are a bit disappointing from an engineering expectation of predictive simulation, they are nevertheless helpful in ascertaining unambiguously the predictions of a given set of physical models in an actual device context. Even better physical models are required to close the gap between experiment and theory, an assertion made possible by using device modeling as a probe of the underlying physical models.

3.3 The importance of dynamic screening in electron-electron scattering

A Monte Carlo study of sub-bandgap impact ionization in short channel silicon-on-insulator (SOI) nMOSFETs [29] unscores nicely the importance of dynamic screening in electron-electron scattering, and in so doing, the general importance of incorporating relevant and accurate physical models in device modeling programs. The background of the physical problem studied is this: SOI devices are thought to have the potential to outperform bulk Si MOSFETs due mainly to reduced parasitic capacitances owing to a SiO_2 buried layer beneath channel and source/drain regions of the device. However, this buried SiO_2 layer no longer permits the potential of the active Si layer between source and drain to be defined via a conventional substrate contact. This "floating substrate" can cause a number of problems for the device designer. One is the so-called "kink effect," where the potential of the floating substrate changes as the drain voltage is increased, which causes a change in threshold voltage, which results in an unexpected rise, or "kink," in the device drain current. The potential rise (potential energy decrease) occurs due to the creation of excess holes in the floating substrate, which raises the substrate potential and decreases the threshold voltage. These holes are created by impact ionization: as electrons move from source to drain, they gain enough energy to impact ionize, creating an electron-hole pair. The electron is collected at the drain, but the hole becomes "trapped" in the potential well of the floating substrate. So far, this behavior is well understood. However, this kink effect has been observed at drain voltages well below 1.12 V, that is, well below the energy gap potential E_{gap}/e for Si. Since E_{gap} is the minimum energy for an electron to impact ionize, how can significant impact ionization occur at drain voltages down to 0.6 V, well below E_{gap}/e? DAMOCLES™ simulations indicate the kink effect will *not* occur at such voltages, unless electron-electron scattering is included in the model and dynamic, rather than the more usual static, screening is employed.

The SOI nMOSFET studied to make this conclusion is a 0.25 μm gate length, 0.18 μm metallurgical channel length device with a 0.08 μm thick, uniformly doped ($N_A = 10^{17}$ cm^{-3}) Si active layer. The active layer sits upon a 0.10 μm thick buried SiO_2 layer; the gate insulator layer is 7 nm thick. Biases are as follows: $V_{GS} = 2$ V and $V_{DS} = 0.8$ V. Fig. 12 shows the device

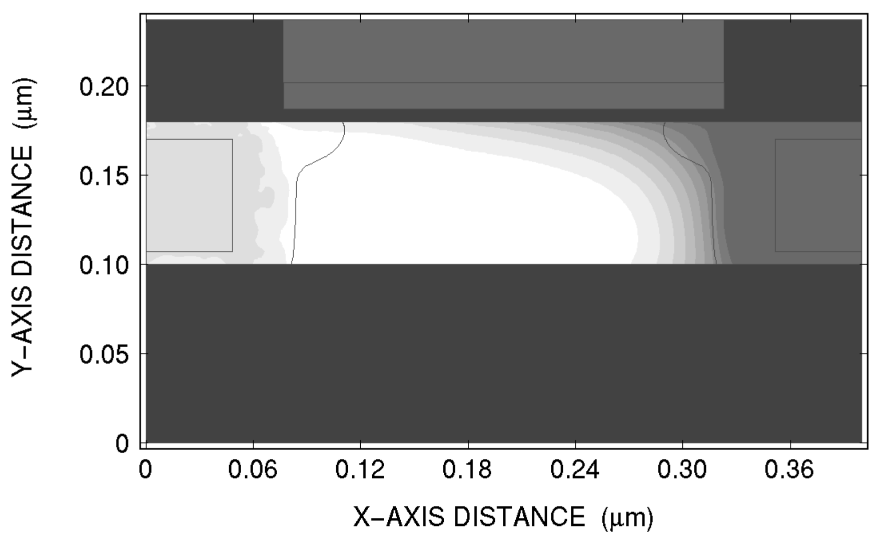

Figure 12: Contours of conduction band energy in the 0.18 μm metallurgical channel length SOI nMOSFET. The darkest gray regions are SiO_2 while the medium gray rectangle centered at the top of the image is the n^+ poly gate. Within the Si layer, potential energy ranges from -0.80 V (middle gray) to $+0.15$ V (white). Here, $V_{GS} = 2$ V and $V_{DS} = 0.8$ V. The metallurgical junction are shown as two curving, vertical lines below the ends of the gate; the remaining lines in the source (left) and drain (right) delimit rectangular regions modeled as contacts.

cross section, and depicts the conduction band energy. The dark gray areas below, and above left and right in the figure, are the SiO_2 layers; the active Si layer is shaded from middle gray to white, corresponding to a potential energy variation from -0.80 to $+0.15$ V (0 V is the Fermi energy in the source contact at the left). Here the contours are seen to crowd near the drain end of the channel, where the electric field is highest. Yet, since $V_{DS} = 0.8$ V, electrons moving *ballistically* only have a kinetic energy of ~ 0.9 eV upon arrival at the source: 0.8 eV gained from the field and ~ 0.1 eV due to their initial energy in the source, obtained because of degeneracy effects there. Fig. 13 shows the

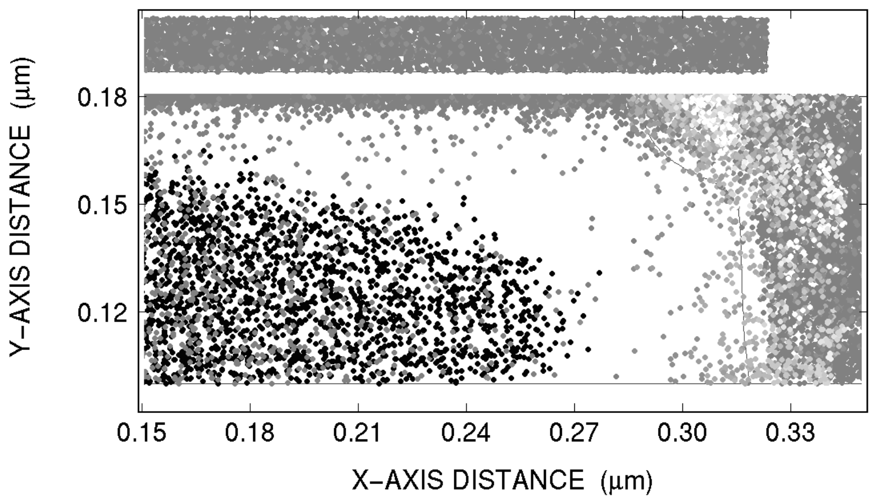

Figure 13: Close up of the drain region of the device of Fig. 12, now showing the location of Monte Carlo particles. Holes are colored black, and are located away from the drain. Electrons are colored according to their kinetic energy: 0 to 1.5 eV spans middle gray to white. Hot electrons are created in the vicinity of the drain, as expected.

device cross section near the drain end of the channel, depicting the position of the Monte Carlo electrons (shaded from middle gray to white as their kinetic energy ranges from 0 up to 1.5 eV), and holes (shaded black). The manner in which holes become trapped in the floating substrate is now quite evident.

Simulations without electron-electron scattering do not predict an electron population near the drain with sufficient energy to impact ionize; hence, no kink effect — at variance with experiments. Including electron-electron interaction opens up the possibility of an energetic electron arriving in the drain and interacting with a drain electron, with a small probability that the energetic electron *gains* energy at the expense of the drain electron losing energy. Yet, the quantitative behavior of this mechanism is *still* unable to establish an electron population near the drain with sufficient energy to impact ionize, when the conventional static screening parameter is used in the electron-electron scattering rate matrix element. This occurs because static screening in the drain is very effective in reducing the strength of the electron-electron interaction: as a hot electron streams into the drain, the cooler drain electrons immediately respond, screening the interaction. However, when dynamic screening in introduced into the device simulation, the picture changes: as a hot electron streams into the drain, the cooler drain electrons are unable to respond — the hot electron's velocity is too great for the sluggish drain electrons to follow. This leaves the electron-electron interaction unscreened, with the net effect that the energetic electrons are now more able to avail themselves to the energy gain mechanism previously described. Fig. 14 shows this final result, in the form

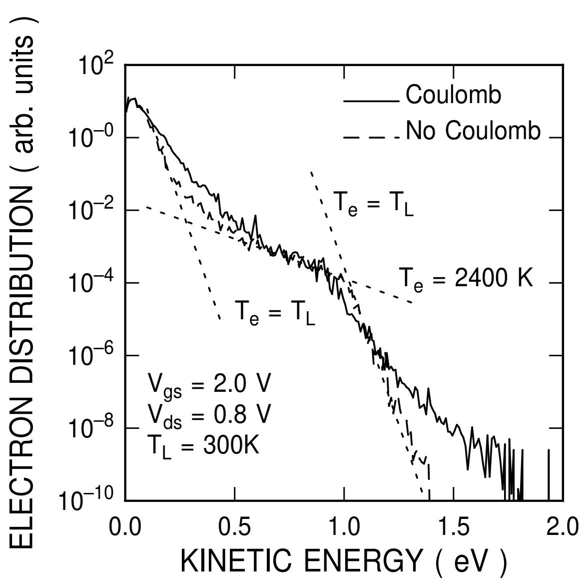

Figure 14: Electron distribution sampled just inside the drain of the SOI nMOSFET simulated. Results with and without the inclusion of Coulomb effects are shown. Full Coulomb effects, including both long- and short-range effects (and the latter modeled with dynamic screening) are necessary to generate a sufficiently energetic electron tail, consistent with the observation of the kink effect at low drain voltages in these devices. Figure adapted from Ref. 29.

of a distribution function of electron kinetic energy. The distribution function samples electrons at the drain end of the channel, and compares the situation with and with Coulomb interactions. Here, Coulomb effects imply both short-range (included via the electron-electron interaction) and long-range (included via the self-consistent solution of the Poisson equation) Coulomb interactions. The difference in the population of energetic (> 1.2 eV) electrons is evident, and finally sufficient in quantity to create enough holes via impact ionization to cause the kink effect, as desired. Thus, the detailed form of the underlying physics is again evident in device characteristics, emphasizing the importance of device modeling in helping to establish the necessary underlying physical models.

4 Conclusions

The relationship between physical models and semiconductor device modeling is both close and symbiotic. Self evidently, the content and accuracy of the underlying physical models determines the quality and predictiveness of the overall device simulation. Less obvious is the opposite influence: determining which physical models are appropriate, and calibrating models, can be expedited by evaluating a model's influence on the same device structure being measured in the laboratory. In so doing, the "computational" laboratory and the actual laboratory are conducting the same experiments, permitting direct comparison of computed and measured device data.

Acknowledgments

One of us (S.E.L.) gratefully acknowledges Prof. Karl Hess for the opportunity of spending a sabbatical with his group at the University of Illinois. DAMOCLES™ is a trademark of the IBM Corporation.

References

1. Semiconductor Industry Association. The national technology roadmap for semiconductors. Technical report, SEMATECH, 2706 Montopolis Dr., Austin, TX, 1994. Also available on the World Wide Web at http://www.sematech.org.
2. *Proc. Int. Workshop on VLSI Processes and Device Modeling (VPAD)*, New York, 1993. IEEE. Held in Nara, Japan, 14-15 May, 1993.
3. *Proc. Int. Workshop on Numerical Modeling of Processes and Devices for Integrated Circuits (NUPAD V)*, New York, 1994. IEEE. Held in Honolulu, HI, 5-6 June, 1994.

4. *Proc. 6th Int. Conf. on Simulation of Semiconductor Devices and Processes (SISDEP '95)*, Vienna, Austria, 1995. Springer-Verlag. Held in Erlangen, Germany, 6-8 September, 1995.
5. *Int. Conf. on Simulation of Semiconductor Processes and Devices (SISPAD '96)*, 1996. To be held in Tokyo, Japan, 2-4 September, 1996.
6. *Proc. 4th Int. Workshop on Computational Electronics (IWCE-4)*, 1995. To appear in *VLSI Design*. Held in Tempe, AZ, 31 October - 2 November, 1995.
7. *5th Int. Workshop on Computational Electronics (IWCE-5)*, 1997. To be held in Notre Dame, IN, 28-30 May, 1997.
8. W. Van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *B. S. T. J.*, XXIX(4):560–607, 1950.
9. *Proc. Computer-Aided Design of IC Processes and Devices*, Stanford, CA, 1996. Stanford University. Held in Stanford, CA, 7-8 August, 1996.
10. J. R. Barker and D. K. Ferry. On the physics and modeling of small semiconductor devices – I. *Solid-State Electronics*, 23(6):519–530, 1980.
11. R. Jancel. *Foundations of Classical and Quantum Statistical Mechanics.* Pergamon Press, Oxford, 1963.
12. C. Kittel. *Quantum Theory of Solids.* John Wiley and Sons, New York, 1963.
13. K. Hess. *Advanced Theory of Semiconductor Devices.* Prentice Hall, Englewood Cliffs, NJ, 1988.
14. U. A. Ranawake, C. Huster, P. M. Lenders, and S. M. Goodnick. PMC 3D: A parallel three-dimensional Monte Carlo semiconductor device simulator. *IEEE Trans. Comput. Aided-Design*, 13(6):712–724, 1994.
15. S. M. Goodnick, S. S. Pinnathur, U. A. Ranawake, P. M. Lenders, and V. K. Tripathi. Parallel implementation of a Monte Carlo particle simulation coupled to Maxwell's equations. *Int. J. Numer. Model., Electron. Netw. Devices Fields*, 8(3–4):205–219, 1995.
16. H. Sheng, R. Guerrieri, and A. L. Sangiovanni-Vincentelli. Massively parallel computation for three-dimensional Monte Carlo device simulation. In W. Fichtner and D. Aemmer, editors, *Simulation of Semiconductor Devices and Processes , Vol. 4*, pages 285–290. Hartung-Gorre Verlag, Konstanz, Switzerland, 1991.
17. W. Kohn and J. M. Luttinger. Quantum theory of electrical transport phenomena. *Phys. Rev.*, 108(3):590–611, 1957.
18. L. Van Hove. Quantum-mechanical perturbations giving rise to a statistical transport equation. *Physica*, XXI(6):517–540, 1955.
19. S. E. Laux and M. V. Fischetti. Monte Carlo simulation of Si n-MOSFETs at 77 and 300 K. *IEEE Electron Device Lett.*, 9(9):467–469,

1988.
20. 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*, 38(14):9721–9745, 1988.
21. M. V. Fischetti, S. E. Laux, and D. J. DiMaria. The physics of hot-electron degradation of Si-MOSFETs: Can we understand it? *Applied Surface Science*, 39(1–4):578–596, 1989.
22. S. E. Laux, M. V. Fischetti, and D. J. Frank. Monte Carlo analysis of semiconductor devices: The DAMOCLES program. *IBM J. Res. Develop.*, 34(4):466–494, 1990.
23. S. E. Laux and M. V. Fischetti. The DAMOCLES Monte Carlo device simulation program. In K. Hess, J. P. Leburton, and U. Ravaioli, editors, *Computational Electronics*, pages 87–92. Kluwer Academic, Boston, 1991.
24. 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 Devices*, 38(3):634–649, 1991.
25. 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 Devices*, 38(3):650–660, 1991.
26. S. E. Laux and M. V. Fischetti. Numerical aspects and implementation of the DAMOCLES Monte Carlo device simulation program. In K. Hess, editor, *Monte Carlo Device Simulation: Full Band and Beyond*, pages 1–26. Kluwer Academic, Boston, 1991.
27. M. V. Fischetti and S. E. Laux. Monte Carlo study of electron transport in silicon inversion layers. *Phys. Rev. B*, 48(4):2244–2274, 1993.
28. M. V. Fischetti, S. E. Laux, and E. Crabbé. Understanding hot-electron transport in silicon devices: Is there a shortcut? *J. Appl. Phys.*, 78(2):1058–1087, 1995.
29. M. V. Fischetti and S. E. Laux. Monte Carlo study of sub-band-gap impact ionization in silicon field-effect transistors. *Proc. Int. Electron Devices Meeting*, pages 305–308, 1995.
30. M. V. Fischetti and S. E. Laux. Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys. *J. Appl. Phys.*, 80(4):2234–2252, 1996.
31. M. V. Fischetti and S. E. Laux. Monte Carlo simulation of electron transport in Si: The first 20 years. To appear, Proc. ESSDERC, Bologna, Italy, 1996. Also available on the World Wide Web

at http://www.research.ibm.com:8080/.

32. M. V. Fischetti, N. Sano, S. E. Laux, and K. Natori. Full band structure theory of high-field transport and impact ionization of electrons and holes in Ge, Si, GaAs, InAs, and $In_{0.53}Ga_{0.47}As$. Submitted to J. Appl. Phys. Also available on the World Wide Web at http://www.research.ibm.com:8080/.
33. Carlo Jacoboni and Lino Reggiani. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Rev. Mod. Phys.*, 55(3):645–705, 1983.
34. J. Y. Tang and Karl Hess. Impact ionization of electrons in silicon (steady state). *J. Appl. Phys.*, 54(9):5139–5144, 1983.
35. Marvin L. Cohen and T. K. Bergstresser. Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures. *Phys. Rev.*, 141(2):789–796, 1966.
36. James R. Chelikowsky and Marvin L. Cohen. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. *Phys. Rev. B*, 14(2):556–582, 1976.
37. George Sai-Halasz, Matthew R. Wordeman, D. P. Kern, S. Rishton, and E. Ganin. High transconductance and velocity overshoot in NMOS devices at the 0.1-μm gate-length level. *IEEE Electron Device Lett.*, 9(9):464–466, 1988.
38. Antonio Abramo, L. Baudry, Rosella Brunetti, Rene Castagne, M. Charef, F. Dessenne, Philippe Dollfus, Robert Dutton, W. L. Engl, R. Fauquembergue, Claudio Fiegna, Massimo V. Fischetti, Sylvie Galdin, Neil Goldsman, Michael Hackel, Chihiro Hamaguchi, Karl Hess, Ken Hennacy, Patrice Hesto, Jack M. Higman, Takahiro Iizuka, C. Jungemann, Y. Kamakura, Hans Kosina, T. Kunikiyo, Steven E. Laux, Hongchin Lin, Christine Maziar, Hiroyuki Mizuno, H. J. Peifer, Sridhar Ramaswamy, Nobuyuki Sano, Paul G. Scrobohaci, Siegfried Selberherr, M. Takenaka, Ting wei Tang, Kenji Taniguchi, J. L. Thobel, R. Thoma, Kazutaka Tomizawa, Masaaki Tomizawa, Thomas Vogelsang, Shiuh-Luen Wang, Xiaolin Wang, Chiang-Sheng Yao, P. D. Yoder, and Akira Yoshii. A comparison of numerical solutions of the Boltzmann transport equation for high-energy electron transport silicon. *IEEE Trans. Electron Devices*, 41(9):1646–1654, 1994.
39. J. Bardeen and W. Shockley. Deformation potentials and mobilities in non-polar crystals. *Phys. Rev.*, 80(1):72–80, 1950.
40. P. D. Yoder, V. D. Natoli, and Richard M. Martin. *Ab Initio* analysis of the electron-phonon interaction in silicon. *J. Appl. Phys.*, 73(9):4378–4383, 1993.

41. M. V. Fischetti and J. M. Higman. Theory and calculation of the deformation potential electron-phonon scattering rates in semiconductors. In K. Hess, editor, *Monte Carlo Device Simulation: Full Band and Beyond*, pages 123–160. Kluwer Academic, Boston, 1991.
42. E. Cartier, M. V. Fischetti, E. A. Eklund, and F. R. McFeely. Impact ionization in silicon. *Appl. Phys. Lett.*, 62(25):3339–3341, 1993.
43. T. H. Ning, C. M. Osburn, and H. N. Yu. Emission probability of hot electrons from silicon into silicon dioxide. *J. Appl. Phys.*, 48(1):286–293, 1977.
44. Z. A. Weinberg. Tunneling of electrons from Si into thermally grown SiO_2. *Solid-State Electron.*, 20(1):11–18, 1977.
45. A. Hartstein, Z. A. Weinberg, and D. J. DiMaria. Experimental test of the quantum-mechanical image-force theory. *Phys. Rev. B.*, 25(12):7174–7182, 1982.
46. G. Binnig, N. Garcia, H. Rohrer, J. M. Soler, and F. Flores. Electron-metal-surface interaction potential with vacuum tunneling: Observation of the image force. *Phys. Rev. B*, 30(8):4816–4818, 1984.