Ultrasmall Devices: Are We Ready for Quantum Effects?

D. K. Ferry

Department of Electrical Engineering and Center for Solid State Electronics Research
Arizona State University, Tempe 85287-5706, ferry@asu.edu

ABSTRACT

It is clear that continued scaling of semiconductor devices will bring us to a regime with gate lengths of less than 70 nm within another decade. While there are questions to be answered in the fabrication of these devices, the questions that must be addressed in simulation are no less difficult. Indeed, pushing to dimensional sizes such as this will probe the transition from classical to quantum transport, and many new issues will arise that must be addressed. In this paper, several of these issues, connected with the discreteness of point charges and with the onset of quantum effects will be discussed.

Keywords: ultra-submicron devices, quantum effects, discrete impurities, molecular dynamics.

1 INTRODUCTION

As the density of integrated circuits continues to increase, there is a need to shrink the dimensions of the devices of which they are comprised. Smaller circuit dimensions leads to more transistors on a single die without negatively impacting the cost of die manufacturing. For silicon MOSFETs, the device size is scaled in all dimensions, resulting in smaller oxide thickness, junction depth, channel length, channel width, and isolation spacing. Advances in lithography have driven device dimensions to the deep-submicrometer range. Currently, 0.18 μm is the state-of-the-art process technology, but even smaller dimensions are expected in the near future. The Semiconductor Industry Association (*SIA*) projects that, by the end of 2009, the leading edge devices will employ 0.05 μm gate lengths and have oxide thickness of 1.5 nm, or less. The group from Toshiba, who has fabricated *n*-channel MOSFETs with effective gate lengths as short as 27 nm [1,2], demonstrated that these feature sizes are feasible.

In simulations of these ultrasmall semiconductor devices, a number of important considerations have been either ignored or have been approximated in a manner which is not representative of the actual physical interactions within the devices. One of these is the Coulomb interaction between the electrons and the impurities and between the individual electrons themselves. This Coulomb interaction has two parts: first, the nature of discrete impurities and how this affects device performance, and, secondly, how the Coulomb interaction affects the *transport* of the carriers through the device. Perhaps of even more importance is the role of quantum effects that begin to occur as the dimensions are reduced to the 50 nm scale.

These effects become even more important when it is realized that simulations must be carried out in the three-dimensional regime. Where simple two-dimensional simulations could suffice in today's arena, it will not be adequate for the 50 nm devices of the future, and it is important to have full three-dimensional simulations if the effects discussed here are to be adequately examined. In the following, I discuss a number of important points which need to be modeled in future efforts, with particular emphasis on those that are of a quantum nature.

2 DISCRETE IMPURITIES

The importance of discrete impurities, and the fluctuations in real impurity density around the nominal doping, were first discussed by Keyes in 1972 [3]. At that time, the concern was variation in threshold voltage that would arise as the fluctuation in dopant number became a significant fraction of the dopant number itself. Wong and Taur [4] studied small Si MOSFETs, and further work showed the effect in GaAs devices [5]. As expected, the presence of a random distribution of dopants in the device, instead of a uniform and constant doping density, led to large variation in threshold voltage and device characteristics.

However, it is important to recognize that the introduction of discrete impurities is not the entire story. Rather, this approach allows one to begin to address more complicated properties of the impurities themselves. That is, it is known from a variety of studies that the impurities do not often site themselves according to the nice average distribution that is usually assumed in the above studies. Instead, there are processing induced effects, such as the tendency of dopants (particularly Boron, linked to both Si and B interstitials [6,7]) to cluster. These lead to impurity distributions different from the smooth one normally assumed. This is complicated by the tendency of other impurities to experience transient enhanced diffusion during anneal after implantation. While these effects have been studied on their own, there is to date no incorporation of such impurity clustering within a semiconductor device model. While of significant importance in Si, impurity clustering and/or ordering is also thought to occur in other materials.

2.1 Coulomb Interactions

Most ensemble Monte Carlo (EMC) simulation of small semiconductor devices include the details of the Coulomb interactions only through a **k**-space scattering process. In general, this process does not account for significant energy loss (or gain) by an individual carrier [8]. In such simulations, Ravaioli [9] has shown that the carriers will go several tens of nm into the drain before relaxing their energy and directed momentum. The problem is that the carriers in the channel interact, through the Coulomb interaction, not only with impurities and other carriers in the channel, but also with impurities and carriers in the drain and gate [10]. The energy loss mechanism of energetic carriers moving from the channel to the drain is by the emission of plasmons (coupled modes of the drain electrons) [10-13]. The real Coulomb interaction is long range, though. Hence, it is important to treat the full Coulomb interaction properly in real space.

Previously, we have discussed a full three-dimensional model of an ultrasmall MOSFET, in which the transport is treated by a coupled EMC and molecular dynamics (MD) procedure to treat the Coulomb interaction in real space [14,15]. The inclusion of the proper Coulomb interaction significantly affects both the energy and momentum relaxation processes, but also has a dramatic effect on the characteristic curves of the device. Relaxation occurs in the drain over a few nanometers, and the Coulomb "scattering" causes a significant shift in threshold voltage as well as a reduction in actual drain current. It is important to note that, within our scheme, we have taken special care to avoid double counting of the Coulomb interaction, as the long range portion is automatically incorporated into the Poisson solution for the self-consistent potential. Only the short range interaction (which is lost by the discretization used for the device simulation) is treated by the molecular dynamics [14]. This real space approach, while computationally intensive, automatically incorporates the full Coulomb interaction. Also, the energy loss that would arise from plasmon emission is fully incorporated with no approximations to the screening.

2.2 Multiple Scattering

Of significant importance is the role that the random distribution (and the full Coulomb interaction) plays on the transport of the carriers in the channel. In studies of transport in bulk systems, in which the dopants have been inserted as discrete items, and detailed Coulomb interactions have been incorporated into an EMC simulation, the mobility is much below what one would estimate from any usual scattering rate based upon the electrons interacting with a single impurity at a time [16,17]. In fact, it was found that the carrier was interacting with several impurities at any given instant of time. This raises a new complication, which is further compounded by any clustering of impurities within the device.

Scattering theory, as it has been developed over the years through the Boltzmann equation, treats each scattering event as an independent quantity so that each scattering event is *fully completed* before the next one is started. Only in this way, can the scattering processes be treated via simple perturbation theory, and the individual processes summed via Mathiessen's rule [18]. When an individual carrier is interacting with several impurities, or other carriers, at the same instant of time, this approach fails. Yet, the simulations, in which the full Coulomb interaction is maintained within the transport model via a molecular dynamics implementation, have shown that this is indeed the situation at high doping densities [16,17]. It should be pointed out that this is not a problem for phonon scattering, as the mean time between collisions is much larger than the duration of any one collision (the latter is of the order of a few femtoseconds [19]). It is a problem of the long-range (and long-time) behavior of the Coulomb interaction between the carrier and other charge centers, and one does have to worry about phonon scattering during a Coulomb interaction (this is of course part of the screening problem that must be addressed in **k**-space implementations).

In a sense, the multiple scattering problem is one of strong interactions in the full many-body system. This means that one should consider a quantum approach (which will be discussed below), but this is not any better developed! Nevertheless, the presence of impurity clusters and the multiple-scattering interactions means that one must include more detailed corrections even to the solutions of the Poisson equation. These additional terms include estimates of the exchange interaction through a linear-density approximation, which provides an additive correction to the potential solutions found from the Poisson equation. These are typically taken to vary as $n^{1/3}$ in three-dimensional systems and provide a general lowering of the energy of the conduction electrons [20]. In fact, these high doping effects are a major contributor to band gap narrowing. We remark that the exchange correction also affects the molecular dynamics through a modification of the Coulomb potential for electrons of the same spin [21]. At the same time, the interaction of the channel electrons with the gate electrons is only now being investigated [10], and these may provide further (important) effects that modify the device behavior at small dimensions.

3. CONDUCTANCE CORRELATIONS

An important aspect of transport theory, as applied to large devices, is the assumption that all correlations of the carriers in the channel with the source are destroyed

immediately upon entering the channel. It is this assumption that allows us to compute the conductance from the non-equilibrium distribution function. In small devices, this assumption is likely to be seriously in error.

Conductance is generally given by the current-current correlation function through the Kubo formula [22]. In strongly degenerate systems, where the velocity is the Fermi velocity, this reduces to the density-density correlation function, quite the same as used in screening of a Coulomb charge. In general, the approach is fully the same as general kinetic theory using the distribution function provided that the correlations are properly included in the latter. We expect the correlations to become important when the device length becomes smaller than, or comparable to, the mean-free path for phase breaking processes (simple scattering doesn't achieve the breakup of the correlations unless it destroys the phase coherence of the carriers). In general, there is no well-developed theory for non-equilibrium transport when the full current-current correlation function must be retained. Instead, nearly all approaches have utilized the kinetic approach, with no thought being given to the correlation of the carriers. Yet, in devices with gate lengths below 50-60 nm, it is not clear that this is a valid approach. It is known at low temperatures, that additional correlated processes can lead e.g. to conductance fluctuations and the break-down of ensemble averaging. When an individual device may contain only a few hundred electrons (and/or impurities), we cannot guarantee ensemble averaging within a single device. In fact, it was just this failure that led Keyes to first consider doping variations as a source of threshold voltage variations across a chip [3].

The conductance correlations are represented as a two-particle interaction, and the reduction to a simple form of the density-density correlation function in Fermi systems is an approximation. This becomes much more important in light of the findings that a carrier in the channel may be interacting with multiple impurities at one time [16,17], as it is just this type of interaction which leads to complications in evaluating the correlation function. We return to this point below.

4. QUANTUM EFFECTS—ARE WE READY?

Generally, when dealing with the quantum transport in a hypothetical device, it is assumed that the electron is represented by an initial wave packet which arises in the source and exits in the drain. Yet this over-constrains the wave function (which must satisfy the Schrödinger equation—a single initial condition in time is the only allowed temporal constraint). In trying to arrive at a proper wave function, a different approach is required in which the wave function is required to be variationally minimized subject to approaching *two positions*, one in the source and one in the drain, in a time-independent manner [23]. This introduces the *quantum localization* problem to device physics—just how small can the two regions be in this formulation (that is, in how small a region can the electron be considered to exist)? In recent work, we have discussed the arguments for various sizes for electrons in semiconductor devices [24]. In particular, in the case of a thermal distribution of electrons, this minimum packet is approximately 60% of the thermal de Broglie wavelength, or about 5 nm at room temperature.

The localized wave packet representation of the carriers then may be associated with an "effective" potential in which sharp discontinuities are removed from the potential within the device [25]. The lack of sharp potentials, and their effective replacement by smoothed potentials, removes much of the driving force for quantization within these device structures. The introduction of such an effective potential leads to an enhancement of the drain-induced barrier lowering in ultra-small devices [26].

This becomes of further importance in considering direct tunneling of electrons from the source to the drain through the above potential barrier. Lowering this barrier by the effective potential makes this a more likely possibility.

4.1 Green's Functions

It has become popular to study small devices through the use of the non-equilibrium Green's function (NEGF). In principle, this approach should build in the full quantum mechanics of the device under study, but this is true only so long as all of the correct interactions are *properly* incorporated into the NEGF itself. One major advantage of this approach is the automatic incorporation of correlation effects in the solution, provided that they are set up in NEGF. What do I mean by this?

The general approach uses a kinetic equation (the quantum equivalent of the Boltzmann equation) to determine the NEGF self-consistently with the potential in the active device [27]. The problem is that a general self-energy is usually assumed for the scattering. This self-energy is developed from the full two-particle interaction Green's function using the Langreth theorem [27], but this theorem is only proven for near equilibrium systems. It uses slow evolution from the equilibrium state to develop a separation of the two-particle Green's function into the product of the NEGF and a self-energy term. In contrast, it is well known in semiconductor devices, where the system is quite far from equilibrium, that the distribution function is a balance between the driving fields and the collisional forces. That is, the far-from-equilibrium state is not guaranteed to be achievable from the equilibrium one by a simple perturbation approach. As a consequence, *there is no first principles proof that this approach is even usable*

for small devices, particularly where transient effects and correlation through the device are likely to be quite important. It is necessary that this separation of the two particle (interaction) Green's function be established. Lacking this proof, the full Bethe-Salpeter equation for this interaction Green's function needs to be investigated thoroughly.

4.2 Multiple Scattering

As an example, let us return to the problem of an electron in the channel interacting with multiple impurities (or other electrons) at any instant of time. To date, this problem has not been addressed either in the self-energy formulation, or in the conductance correlation function (itself a two-particle Green's function). In theory, one set of multiple scattering diagrams (the nested diagrams) can be treated through a vertex correction to the impurity (or Coulomb) scattering interaction. Such a vertex correction was considered at low temperatures as impurity screening of the electron-electron interaction [28,29]. There has been no consideration of these terms in device modeling.

It is clear that a first principles investigation of the foundations of multiple scattering needs to be done. This can be done either within the NEGF formulation, or within the Bethe-Salpeter equation for the conductance correlation function. Without such an investigation, the proper understanding of quantum effects within the ultra-submicron device cannot be achieved.

V. CONCLUSIONS

We have only begun to uncover the important effects that will affect device modeling for structures with critical lengths of 50 nm and below. There remains hard work to incorporate the results of high density and doping, and to develop a proper quantum mechanical treatment of device modeling.

This work is supported by the Semiconductor Research Corporation, the Office of Naval Research, and the National Science Foudnation.

REFERENCES

[1] M. Ono, M. Saito, T. Yoshitomi, C. Fiegna, T. Ohguro and H. Iwai, IEDM Tech. Dig. 119, 1993.

[2] M. Ono, M. Saito, T. Yoshitomi, C. Fiegna, T. Ohguro and H. Iwai, IEEE Trans. Electron Devices 42, 1822, 1995.

[3] R. Keyes, Science 195, 1230, 1975.

[4] H.-S. Wong and Y. Taur, IEDM Tech. Dig. 705, 1993.

[5] J.-R. Zhou and D.K. Ferry, IEEE Comput. Science and Eng. 2, 30, 1995.

[6] T. E. Haynes, D. J. Eaglesham, P. A. Stolk, H.-J. Gossmann, D. C. Jacobson, and J. M. Poate, Appl. Phys. Lett. 69, 1376, 1996.

[7] J. Zhu, T. Diaz de la Rubia, L. H Yang, C. Mailhiot, and G. H. Gilmer, Phys. Rev. B 54, 4741, 1996.

[8] D. K. Ferry, S. M. Goodnick, and K. Hess, Physica B 272, 538, 1999.

[9] U. Ravaioli, *private communication*.

[10] M. V. Fischetti and S. E. Laux, *submitted for publication*.

[11] M. A. Hollis, S. C. Palmateer, L. F. Eastman, H. V. Dandekar, and P. M. Smith, IEEE Electron Dev. Lett. 4, 440, 1983.

[12] P. Lugli and D. K. Ferry, IEEE Electron Dev. Lett. 6, 25, 1985.

[13] P. Lugli and D. K. Ferry, Appl. Phys. Lett. 46, 594, 1985.

[14] W. J. Gross, D. Vasileska, and D. K. Ferry, IEEE Electron Dev. Lett. 20, 463, 1999.

[15] W. J. Gross, D. Vasileska, and D. K. Ferry, IEEE Trans. Electron Dev., *submitted for publication*.

[16] R. P. Joshi, and D. K. Ferry, Phys. Rev. B 43, 9734, 1991.

[17] D. Vasileska, W. J. Gross and D. K. Ferry, Proc. Intern. Workshop on Computational Electronics, Kyoto, 1998.

[18] D. K. Ferry, "Semiconductors," Macmillan, 1991.

[19] P. Bordone, D. Vasileska, and D. K. Ferry, Phys. Rev. B 53, 3846, 1996.

[20] D. Vasileska and D. K. Ferry, IEEE Trans. Electron Dev. 44, 577, 1997.

[21] A. M. Kriman, M. J. Kann, D. K. Ferry, and R. Joshi, Phys. Rev. Lett. 65, 1619, 1990.

[22] R. Kubo, J. Phys. Soc. Jpn. 12, 570, 1957.

[23] See, e.g., the discussion in L. S. Schulman, "Time's Arrow and Quantum Measurement" (Cambridge Univ. Press, 1997), and references therein.

[24] D. K. Ferry and H. L. Grubin, Proc. Intern. Workshop on Computational Electronics, Kyoto, 1998, 84.

[25] A. M. Kriman, J.-R. Zhou, and D. K. Ferry, Phys. Lett. A 138, 8, 1989, and references therein.

[26] D. K. Ferry, Superlatt. Microstruc., *in press*.

[27] See, e.g., D. K. Ferry and H. L. Grubin, Sol. State Phys. 49, 283, 1995.

[28] H. Fukuyama, J. Phys. Soc. Jpn. 49, 644, 1980.

[29] D. K. Ferry and S. M. Goodnick, "Transport in Nanostructures" (Cambridge Univ. Press, 1997) Sec. 6.3.6.