Transport Physics in Semiconductor Nanostructures

David K. FERRY

Center for Solid State Electronics, Arizona State University
Tempe, Arizona 85287-6206 USA

The progress in Si ULSI, and the parallel progress in GaAs LSI, has pushed the gate length in current devices toward 0.1 μm. Research laboratories have gone even further, and FETs with gate lengths as short as 20 nm have been demonstrated. It is apparent that within the next decade or so, devices with such gate lengths will be made in the production environment. These devices open the door to new physical effects that are now being studied in nanostructures, or mesoscopic devices as they are usually referred to. The first new effect, already seen in the short-gate length devices, is tunneling through the gate depletion region. In this talk, I will try to summarize the major physical effects and the role they may play in future industrial devices.

1. INTRODUCTION

Since the introduction of integrated circuits, the number of individual transistors on a single chip has doubled approximately every three years. Today, we are talking about multi-megabit DRAM memories (the 16 Mb is on the market, the 64 Mb is in pre-production, and research versions of the 256 Mb have been demonstrated) and dense signal-processing chips with comparable component density. At the rate of progress of dynamic memory (DRAM), we can expect to reach chip densities of 10^9 devices by 2001. By 2020, we will need to have memory chips with 1 Tb. In general, progress in the integrated circuit field has followed a complicated scaling relationship. The reduction of design rule (or effective gate length) proceeds approximately by a factor of 1.4 each generation (which produces only an increase of 2x in density, the remainder coming from circuit enhancements and larger chip size). This means we will be using 0.1-0.15 μm rules for the 4 Gb chips (the 256 Mb chip will use 0.25 μm design rules). If we continue this extrapolation, current technology will require 30 nm design rules, and a cell size of 670 nm^2, for a 1 Tb memory chip.

It appears then that we will eventually see devices with gate lengths ≤0.05 μm. An electron traveling at the saturated velocity will traverse this length in 0.5 ps, or approximately the transient time at 50 kV/cm (in Si). Moreover, the de Broglie wavelength for the carriers is only about 0.18 μm, or one-third of the gate length. Thus, it is expected that quantum effects will certainly appear in such devices. When devices of 30 nm (or less) gate lengths are made, it is found that their performance is different from that of current FETs. Research devices fabricated both at Arizona State University, and at Sony in Japan, with gate lengths of 25-30 nm, clearly show that tunneling through the gate depletion region is the dominant contributor to current, and gate control is much reduced due to this effect.

The transport of carriers in semiconductor devices has long been a subject of much interest, not only for material evaluation, but also in the realm of device modeling and, more importantly, as an illuminating tool for delving into the physics governing the interaction of electrons (or holes) with their environment. From the above discussion, it appears that more detailed modeling of quantum effects needs to be included in device modeling for future ultrasmall devices. These quantum effects appear in many guises: a) modification of the statistical thermodynamics, b) introduction of new length scales, c) ballistic transport and quantum interference, and d) new fluctuations affecting device performance. Many of these effects have been studied, either in models of ultra-submicron devices, or in transport studies of nanostructures (mesoscopic devices) at low temperature. In this paper, I will try to introduce a number of these effects and discuss the manner in which they may well influence device performance in the ultra-submicron gate length regime.

2. STATISTICAL THERMODYNAMICS

Generally, modeling of quantum phenomena is more complicated than modeling of classical and/or semi-classical phenomena. For instance, one must consider the full non-local nature of the potential interactions in the dynamical variables. Consider a simple potential barrier \( V(x) = V_0 \mu(x) \), where \( \mu(x) \) is the Heavyside step function. We assume that there is some density existing in the region \( x < 0 \). In Fig. 1, we show the Wigner distribution function for \( V_0 \to \infty \). We note that far from the barrier, the distribution approaches the classical Maxwellian form, but near the barrier, it differs greatly. The repulsion from the barrier is required by the vanishing of the wave
function at the barrier, but the first peak away from the barrier (in the wave function) occurs closer to the barrier for higher momentum states. This leads to much of the complication evident in the figure. The overshoot occurs to accommodate the need for total charge neutrality. This variation exists over a distance of the order of several thermal de Broglie wavelengths, \( \lambda_D = (\hbar^2/2m(kT))^1/2\), which provides a spatial scale length. Thus, nonlocal variations can be expected over a range of 10-20 nm even at room temperature!

Figure 1. The Wigner distribution function for an infinite barrier.\(^4\)

It is clear that the density no longer varies simply as \( \exp(-\beta V) \), where \( \beta \) is the inverse temperature, and that modifications to the statistical mechanics need to be made. Unfortunately, there is no consensus as to the form of the quantum potential correction to this simple exponential. One form that has been used\(^2\) introduces a quantum pressure term as a modification of the electron temperature, through

\[
\beta^{-1} = k_B T_e - \frac{\hbar^2}{8m^2} \nabla^2 \ln(n),
\]

although other work has reduced the last term by a factor of \( 3^3 \). Although the results obtained using this model are in agreement with the intuitive expectations, it should be noted that the correction term does not have the momentum dependence expected from Fig. 1. A more recent derivation overcomes this limitation\(^4\), but has not been tested in actual simulations as yet.

3. PHASE INTERFERENCE

The relevant quantity for discussion of quantum interference effects is the phase of the carrier as it moves through the semiconductor. Interference between differing waves can occur over distances on the order of the coherence length of the carrier wave, and the latter distance is the inelastic mean free path, or phase-breaking length. This is related to the energy relaxation length \( l_e = \tau_e c_0 \), where \( \tau_e \) is the energy relaxation time and \( c_0 \) is a characteristic velocity. The inelastic mean free path can be quite long, on the order of several tens of micrometers for electrons at low temperature in the inversion channel of a high-electron-mobility GaAlAs-GaAs heterojunction structure, but is of the order of 0.1 \( \mu \)m even at room temperature. If the phase remains coherent over the range of the pertinent correlation function (in space or time), there are interference effects that induce correlations between the initial and final directions of the velocities. This is the cause of weak localization in which the scattering processes tend to induce back-scattering which results in a negative correlation between initial and final velocities and lowers the conductivity. The full impact of phase interference leads not only to weak localization, but also to the Aharonov-Bohm effect, universal conductance fluctuations, and resonant tunneling. A full treatment of these mesoscopic effects is beyond the scope of this limited introduction, but interested parties are referred to the excellent reviews edited by Altshuler \textit{et al.}\(^5\) and by Namba \textit{et al.}\(^6\).

One direct result of a long inelastic mean free path, and a correspondingly long elastic mean free path is ballistic transport in which an electron follows a long mean flight without undergoing scattering. This distance can be several tens of nanometers, even at room temperature. When this occurs, the gate loses control of the carriers in the channel, since it can deplete the channel only for lengths longer than the scattering length.

If ballistic trajectories go around an obstacle, such as an impurity atom (or its potential), the interference between these trajectories leads to the above-mentioned phase interference. In short the interference arises from the Aharonov-Bohm effect, in which the action around the coupled trajectories 1 and 2 is modulated by a magnetic field. This follows

\[
\Psi_{\text{total}} \sim \exp\left\{ \int \left[ A \cdot dl - \frac{1}{2} (A \cdot dl) \right] \right\}
\]

\[
- \exp\left\{ \frac{i}{\hbar} [\mathbf{B} \cdot ndI] \right\} = \exp(\Phi/\Phi_0)
\]

where \( \Phi_0 = \hbar/2e \) is the flux quantum. Normally, one applies a magnetic field to study this effect. However, a 10 mA current through an interconnect produces a magnetic field of 200 G at a distance of 0.1 \( \mu \)m, so that a coupling area of 0.01 \( \mu \)m\(^2\) will produce a noticeable phase change. Especially during transients, we can expect additional fluctuating \textit{(in current, not in time)} effects in these ultrasmall devices. We will see this in the next section.

4. LATERAL WAVEGUIDING

In the transport of ballistic electrons through quasi-one dimensional wires, it is generally found that if the length of the wire is quite short (almost through point contacts), the conductance shows steps equal to \( 2e^2/h \) times the number of propagating modes. If the wire is longer, however, even when it remains short compared to the inelastic mean free path, the conductance steps are usually lost. This has been explained generally by the assumption that random potentials induce a dissipative phase loss even though the inelastic mean free path is usually much larger than the wire length. We have investigated this with a waveguide modal approach to the quasi-one dimensional wire, in which we randomly vary the width of the wire. The various inhomogeneous sections lead
to extensive interference effects and the average conductance (over many equivalent wire samples) shows a general disappearance of the conductance steps, even though the inelastic mean free path remains constrained to the wire length (by the computational method). Thus, we conclude that it is the conductance averaging in the inhomogeneous wire, rather than a dissipative phase loss mechanism that leads to the loss of the conductance steps. Thus, while one can expect to see conductance quantization due to lateral confinement effects, the random potentials that will be imposed upon the active region due to fluctuations in its size (by fabrication) and to fluctuations in the device-device interactions will cause much of the quantization to be averaged out\(^7\). However, this averaging will introduce much fluctuations in the actual conductance of the active device region and we must learn how to control and reduce these fluctuations. An example is shown in Fig. 2, where the conductance is plotted as a function of the Fermi energy.

![Figure 2. The conductance of a short wire with a random width. The two curves are from a width fluctuation that is a constant fraction of the Fermi wavelength (top) or a constant fraction of the total width (bottom).](image)

5. GRANULARITY EFFECTS

If we consider a gate area of $30 \times 30 \text{mm}^2$, then a gate oxide of 5 nm leads to a capacitance of 0.6 aF. While one might think this thin oxide will be a tunneling oxide, an electron cannot tunnel through the oxide unless the gate voltage is $\pm 0.26$ V (which is 10 $k_BT$). This feature is known as Coulomb blockade, in which the voltage must be large enough to provide the energy change required by the tunneling electron. Single-electron charging phenomena, which cannot be accounted for without considering the quantization of charge, have recently attracted much attention. In metal tunnel junction systems, new physical phenomena, which include SET oscillation and the Coulomb staircase, have been found. Many experimental and theoretical efforts have been directed to the underlying physics of this phenomena\(^8\). Single-electron charging phenomena are observed not only in metal systems but also semiconductors such as Si-MOSFET and GaAs heterojunction 2-DEG systems, and insulators such as InO$_2$. In metal tunnel junctions, the concept of Coulomb blockade and the "orthodox theory," which is a semiclassical treatment using macroscopic capacitances, charges, and potentials, have advanced the understanding of these phenomena by successfully explaining many observed results. These single electron phenomena are thought to have tremendous possibilities in future electronics devices.

In our own simulations, we have sought to gain insight into the above crucial question via numerical simulations\(^9\). We use a classical molecular dynamics technique so that the strong electron-electron correlations in the low-density regime and non-tunneling transfer mechanisms in smooth continuous potentials can be treated. The molecular dynamics approach to electron transport inherently incorporates the exact microscopic correlations, including multi-electron scattering, without approximation (within the classical regime). This approach successfully reproduces the experimentally observed conductance oscillations, activation energy, and threshold voltage of the nonlinear conductance of a classical double-barrier structure. The simulation also gives numerical evidence that significant single-electron charging effects can be observed in continuous, smooth potentials even without tunneling. This opens the door to a wide variety of single-electron effects in conventional semiconductor devices. Indeed, scientists at Hitachi have made a multiple-well single-electron tunneling transistor in Si MOS technology that closely resembles our computed structure. Operation in the SET mode was observed easily at 4.2 K, even though the geometries were quite relaxed (well diameters of approximately 0.15 $\mu$m)\(^10\). Work at Hitachi’s Cambridge Research Laboratory has proposed many novel logic circuits with single-electron transistors.

6. REFERENCES

8. See, e.g., the article by D. V. Averin and K. K. Likharev in ref. 5.