# Density-Gradient Analysis of Tunneling in MOS Structures with Ultra-Thin Oxides

M.G. Ancona
Naval Research Lab, Washington, DC 20375
ancona@sinh.nrl.navy.mil

Z. Yu and R.W. Dutton
Stanford University, Stanford, CA 94305

P.J. Vande Voorde, M. Cao and D. Vook
Hewlett-Packard, Palo Alto, CA 94306

**Abstract--Quantum transport theory in the density-gradient approximation is applied to the analysis of tunneling phenomena in ultra-thin oxide (<25Å) MOS structures. Detailed comparisons are made with experimental I-V data for samples with both $n^+$ and $p^+$ polysilicon gates and all of the features of this data are found to be understandable within the density-gradient framework. Besides providing new understanding of the experiments, these results show the density-gradient approach to be useful for engineering-oriented device analysis in quantum regimes with current flow.**

## I. INTRODUCTION

Advanced silicon technology is exploiting ever thinner layers of $SiO_2$ and it is becoming increasingly important to be able to understand, model and engineer the quantum mechanical effects associated with these thin layers. For example, inversion layer confinement effects become strongly manifested in this regime and for this reason have been widely studied [1]. Equally important are effects associated with quantum mechanical tunneling. Such tunneling has of course long been central to floating-gate memories, however, more recently it has become a critical issue in ultra-scaled MOSFETs [2]. In an extreme example [3] gate oxides were made as thin as 9Å and, although the devices performed supremely well ($g_m$=1.2S/mm), their electrical characteristics, stand-by power and perhaps long-term reliability were adversely affected by their large gate tunneling currents. Heretofore effects of this type have been modeled with microscopic approaches that are either overly simple [4] or too complex for broad engineering application [5]. This situation led us to develop density-gradient theory [6] which we here apply to ultra-thin MOS tunneling. By making close comparisons with experimental I-V data we both validate the density-gradient approach and improve our understanding of the ultra-thin oxide MOS physics.

## II. DENSITY-GRADIENT THEORY

Density-gradient (DG) theory is a macroscopic approximation to quantum mechanics in which the nonlocality of quantum mechanics enters the theory solely via an assumption that the equations of state of the electron and hole gases depend not only on the gas densities as in diffusion-drift theory but also on the *gradients* of their densities. More microscopically, this approach may be viewed as retaining only the lowest order terms in a density-gradient expansion of full quantum transport theory [7]. In previous work [8] we showed that with this simple foundation a theory is obtained that is often surprisingly accurate in describing effects of quantum confinement and tunneling. And because of its simplicity, DG theory can more readily incorporate and efficiently deal with other complications of real devices such as multi-dimensionality that are not so easily treated by alternative microscopic theories. Moreover, as a generalization of the conventional diffusion-drift description, DG theory fits naturally into the framework of conventional device simulation. Steps toward exploiting this latter advantage have been undertaken in Ref. 9 where density-gradient corrections have been incorporated into the general-purpose simulation code PROPHET from Lucent Technologies.

The key equation of DG theory (for electrons) [5] is an expression of Newton's Second Law which in 1-D takes the form

$$\alpha_n \frac{dv_n}{dt} = (\psi - \varphi_n)_x - \frac{v_n}{\mu_n}$$

where

$$\varphi_n = \varphi_n^o(n) - \frac{2b_n}{s} s_{xx} \quad , \qquad s = \sqrt{n}$$

$\alpha_n$ is the electron mass-to-charge ratio, $b_n$ measures the strength of the gradient effect in the electron gas [7], $\varphi_n$ is the chemical potential, $v_n$ is the average electron velocity and all other parameters have their usual meanings. A variety of microscopic derivations for the coefficient $b_n$ have been given [8]

and their results may be summarized as

$$b_n = \frac{\hbar^2}{4m_n^* q r_n}$$

where we take $r_n$, a dimensionless parameter accounting for statistical effects, to have its pure state value of one [7,8]. In semiconductor regions the scattering ($\mu_n$ term) dominates the inertia ($\alpha_n$ term) and the latter can be neglected. However, in barrier regions the tunneling is typically elastic meaning the mobility is infinite and in this case the inertia term must be non-negligible (at least in some portion of the barrier) as it ensures that the electron velocity remains finite. Furthermore, without scattering to mix the electrons the electron population in the barrier splits in two with each subpopulation having the electrochemical potential of its originating electrode. Since we include holes in the calculations also this means we have *four* distinct carrier types in the barrier region. Each is governed by its own transport equation (continuity plus momentum balance) and, in addition, each is coupled via the electrostatics and must satisfy a set of consistent boundary conditions.

## III. APPROACH

In this paper we use DG theory to model MOS tunneling measurements. This serves both to evaluate DG theory and to provide better understanding of the observed device characteristics. The experimental data were obtained on well-characterized MOS capacitors and MOS test transistors (operated as gated diodes) with $n^+$- and $p^+$-poly gates and with oxides in the range of 11-40Å. These devices were fabricated and electrically tested at Hewlett-Packard's ULSI Research Laboratory. The DG theory "predictions" are obtained by formulating an appropriate boundary value problem in DG theory and solving it using numerical methods. Our numerical scheme is a standard fully-coupled Newton approach which we have found to work best when applied to a quasi-Fermi level formulation of the equations. This has been implemented in a general-purpose code which is applicable to arbitrary 1-D heterostructures. Solutions are obtained both for dc and for small-signal ac conditions though only the former are exhibited here. In viewing these solutions --- for I-V curves, band diagrams, etc. --- one should be cognizant of their various uncertainties and the degree to which they may be regarded as true predictions. First, there are the approximations of DG theory for which our earlier work [8] provides some level of trust as do the findings of this paper. Second and more important, the MOS problem contains some imprecisely known quantities relating to the oxide --- tunneling effective mass, barrier height and oxide thickness --- which are crucial since the tunneling currents depend exponentially on them. Indeed, one can say that no tunneling theory is capable of *ab initio* predictions without a detailed model of the oxide. And for this reason, all such simulations including ours contain an element of curve-fitting when matching theory to data. For our work, the best that can be said is that the values we assume for the oxide parameters are physically reasonable and internally consistent.

## IV. SIMULATION RESULTS

In Fig. 1 the band diagram and carrier density profiles as calculated by DG theory are shown for an $n^+$-poly capacitor with $t_{ox}$ = 15Å and $V_g$=-0.7V. As seen in the Figure, barrier penetration by both electrons and holes occurs with the familiar exponential decay. The existence of four carrier types in the barrier with appropriate directionality is evident in the Figure. For this bias significant current is carried across the barrier by both electrons (net flow to the right) and holes (net flow to the left). The actual simulation region was 2μm long thus illustrating the ability of the DG description and our numerical scheme to deal with slowly-varying conventional bulk device behavior while at the same time resolving the rapid variations occuring in the tunnel barrier. This consistent treatment of electron and hole transport in the semiconductors and in the barrier contrasts with many conventional analyses where the semiconductor is treated classically and tunneling is grafted on using formulas of one-electron quantum mechanics [8].

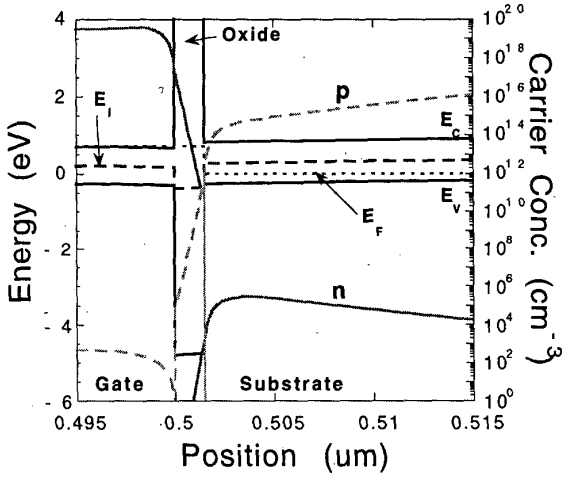


Fig. 1. Band diagram and profiles showing barrier region.

In Fig. 2 the measured I-V characteristics (points) of an $n^+$-poly test transistor with three different oxide thicknesses and with the source grounded (Fig. 2a) or floating (Fig. 2b) are compared with DG simulation results (lines). All cases show excellent agreement. The differences between the two Figures (seen for positive bias) arise because with the source floating the inversion layer is not grounded, the tunnel barrier is in series with the depletion layer and current flow is limited not by tunneling but by the small depletion layer generation.

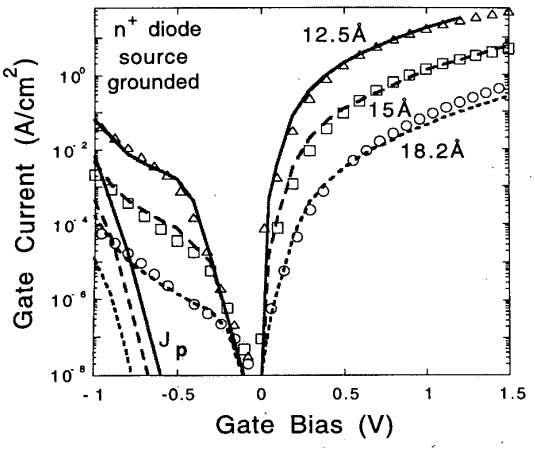


Fig. 2a. Measured (points) and DG simulated (lines) J-V.

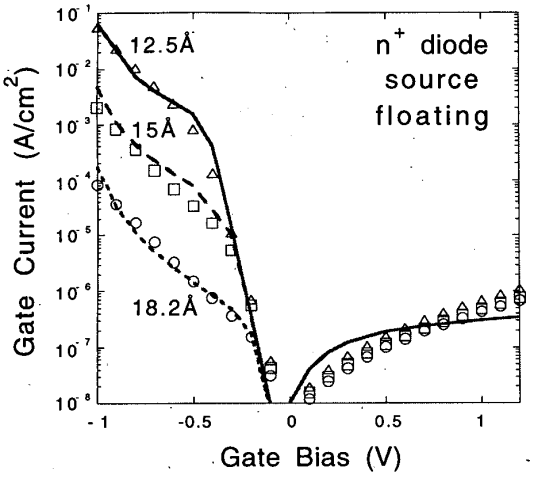


Fig. 2b. Measured (points) and DG simulated (lines) J-V.

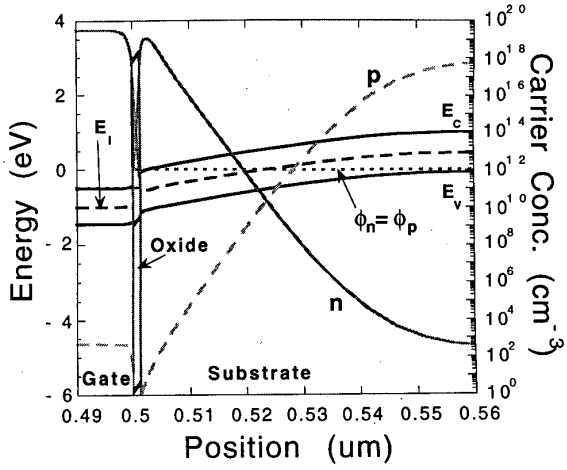


Fig. 3a. Profiles in an n+-poly diode with source grounded.

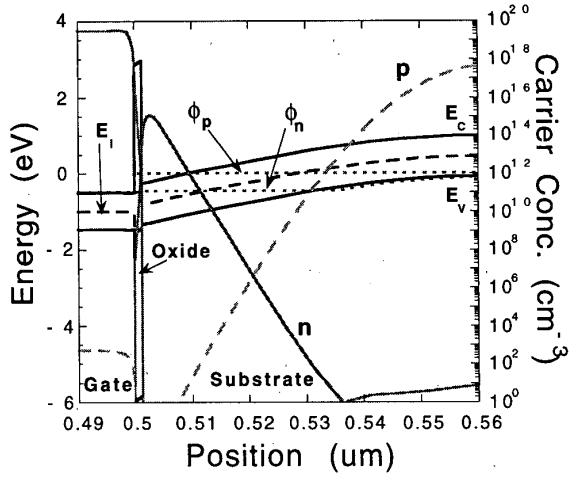


Fig. 3b. Profiles in an n+-poly diode with source floating.

For the latter, in simulation we take $\tau_n = \tau_p = 10^{-8}$ sec so as to get reasonable agreement with the data. This analysis also explains why, with the source floating, the current is independent of the oxide thickness for inversion biases (Fig. 2b, positive bias). The effect of the source contact can also be seen in band diagrams in Fig. 3. With the source floating (Fig. 3b) there is a non-equilibrium variation in the minority carrier quasi-Fermi level that results from the inversion layer being in better contact with the gate (via oxide tunneling) than with the substrate. If the lifetime were reduced sufficiently then the depletion layer generation would grow and eventually short-circuit the tunneling effect causing Figs. 2a and 3a to approach Figs. 2b and 3b.

Another interesting feature of the I-V data and simulation (Fig. 2) is the change in slope that occurs under negative bias. This results from the fact that when the oxide is less than about 25Å the built-in voltage is enough to invert the substrate. When negative bias is applied the bands gradually unbend with most of this unbending occuring in the oxide so long as the inversion layer exists. The changing barrier shape causes the current to increase rapidly. But when the inversion layer suddenly disappears the band-bending in the substrate starts to decrease significantly, the changes to the barrier shape moderate and, as a result, the rate of increase in tunneling current also slows. A plot of the simulated areal electron density in the substrate as a function of the forward bias (Fig. 4) supports this explanation: The break points in the I-V

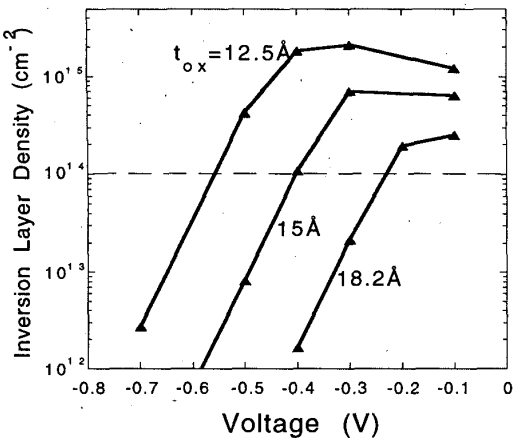


Fig. 4. Inversion layer density in n$^+$-poly transistors with bias.

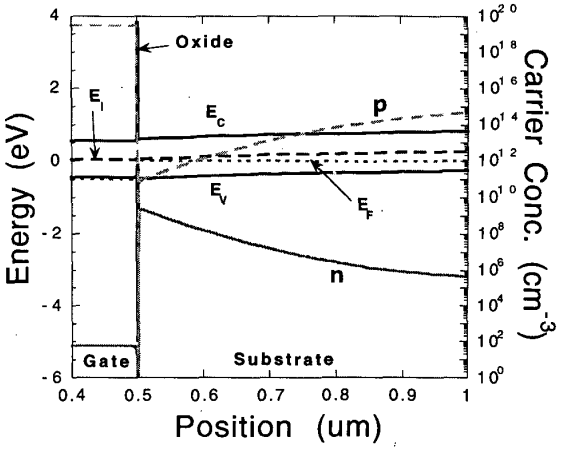


Fig. 6. Band diagram of 17.3Å p$^+$-poly capacitor with V = +0.5V.

curves occur when the inversion layer begins to disappear. As is seen experimentally, the voltage at which this happens is largest for the thinnest oxide.

A final point relating to these simulations may be seen in the simulated I-V curves of Fig. 2a (lower left) where we show the onset of a significant hole tunneling contribution. The source of these holes is the growing accumulation layer in the p-type substrate. In this regime it becomes increasingly important that holes be included in the simulation.

As a last set of comparisons between simulation and experiment in Fig. 5 we exhibit I-V results for p$^+$-poly capacitors (with much lower substrate doping). Good agreement between theory and experiment is again obtained. The saturation under positive bias results from the lightly-doped substrate being depleted at these biases (Fig. 6). This causes the incremental voltage to be dropped across the depletion layer rather than the oxide and so the tunneling current is little affected.

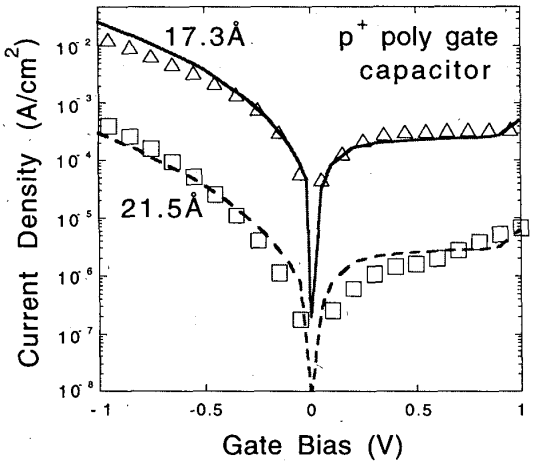


Fig. 5. Measured (points) and DG simulated (lines) J-V.

## SUMMARY

Overall, the DG simulations of this work compare quite favorably with data describing all of the features of the measurements in reasonably quantitative fashion. Besides providing new understanding of these experiments, our results demonstrate that the density-gradient approach can be of great value for engineering-oriented device analysis in quantum regimes particularly when other complications of real devices are important.

## ACKNOWLEDGMENT

We thank B. Biegel and C. Rafferty for useful discussions and MGA thanks ONR for funding support.

## REFERENCES

1. A density-gradient theory treatment appears in M.G. Ancona, Z. Yu, W.-C. Lee, R.W. Dutton and P.J. Vande Voorde, *J. Tech. Comp.-Aided Design* (1998) [http://tcad.stanford.edu/tcad-journal/archive].
2. Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S.-H. Lo, G. Sai-Halasz, R. Viswanathan, H.-J. Wann, S. Wind, H.-S. Wong, *Proc. IEEE* **85**, 486 (1997).
3. G. Timp *et al.*, *IEDM Tech. Dig.*, 930 (1997).
4. See, *for example*, A. Gupta, P. Fang, M. Song, M.-R. Lin, D. Wollesen, K. Chen, C. Hu, *Elect. Dev. Lett.* 18, 580 (1997).
5. C. Bowen, C. Fernando, G. Klimeck, A. Chatterjee, D. Blanks, R. Lake, L. Hu, J. Davis, M. Kulkarni, S. Hattangady, I.-C. Chen, *IEDM Tech. Dig.*, 869 (1997).
6. M.G. Ancona and H.F. Tiersten, *Phys. Rev.* **B35**, 7959 (1987).
7. F. Perrot, *Phys. Rev.* **A20**, 586 (1979); M.G. Ancona & G.J. Iafrate, *Phys. Rev.* **B39**, 9536 (1989).
8. M.G. Ancona, *Superlatt. Microstr.* **7**, 119 (1990); *Phys. Rev.* **B42**, 1222 (1990).
9. C.S. Rafferty, B. Biegel, Z. Yu, M. Ancona, J. Bude and R. Dutton, *Proc. SISPAD* (1998), p. 137.