Modeling of Direct Tunneling Current Through Gate Dielectric Stacks

S. Mudanai, Y-Y Fan, Q. Ouyang, A. F. Tasch, F.Register, D. L. Kwong, and S. Banerjee
Microelectronics Research Center, Dept. of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78758.

Abstract- **The direct tunneling current has been calculated for the first time from an inverted p-substrate through different gate dielectrics by numerically solving Schrodinger's equation and allowing for wavefunction penetration into the gate dielectric stack. The numerical solution adopts a first-order perturbation approach to calculate the lifetime of the quasi-bound states. This approach has been verified to be valid even for extremely thin dielectrics (0.5nm). WKB solution agrees well with the tunneling currents predicted by this technique. For the same effective oxide thickness (EOT) the direct tunneling current decreases with increasing dielectric constant, as expected. However, in order to take full advantage of using high-K dielectrics as gate insulators the interfacial oxide needs to be eliminated. We also present for the first time the CV curves obtained assuming that the wavefunction penetrates into the oxide.**

I. INTRODUCTION

As the gate lengths of CMOS devices continue to scale down to the sub-100nm regime, the device requirements dictate that the gate oxide thickness be scaled down to below 2nm. However, as the thickness decreases below ~2.5nm, the direct tunneling current increases rapidly. In order to decrease the leakage current caused by tunneling, the physical thickness must increase, while the EOT must continue to scale down. This is only possible with the introduction of suitable high-dielectric constant (high-K) materials. Apart from several integration problems [1], device simulations have shown that unforeseen effects come into play when the dielectric constant is increased beyond a certain value due to fringing-induced-barrier-lowering (FIBL) [2,3]. Furthermore, the barrier height to the gate tunneling current generally decreases with increasing dielectric constant. Hence the advantages and disadvantages of choosing different dielectrics must be carefully evaluated.

Recently, an analysis was performed of the dependence of the tunneling currents on the various dielectrics [4]. However, this work did not account for quantization inside the silicon substrate. It also did not take into account the penetration of the wavefunction into the gate dielectric stack. The wavefunction penetration becomes increasingly relevant with scaling down to ultra-thin dielectrics; but very little work has been done to account for the wavefunction penetration inside the oxide [5,6,7]. In [5] a numerical study was done for the modeling of accumulation layers and tunneling currents in thin oxides. However, in this study the wavefunction inside the oxide was assumed to go to zero at the oxide-gate electrode interface. A more general approach would be to assume that the wavefunction penetrates into the oxide and the gate electrode material. In [6,7] a first-order perturbation approach was developed to solve Schrodinger's equation allowing for wavefunction penetration into the gate oxide and gate electrode and compared with the WKB method for oxide thicknesses > 1.5nm. However, the validity of the perturbation approach applied to extremely thin oxides (~0.5nm) was not investigated.

In this work the validity of the previous approach [6] is examined. The numerical approach developed by Shih et. al.[6] has been extended to obtain the wavefunction penetration inside stacked gate dielectrics. The tunneling currents through various dielectrics and dielectric stacks have been calculated in order to obtain an understanding of their dependence on barrier heights and dielectric constants. Finally, the effects of wavefunction penetration through thin dielectrics on quantization in the substrate are investigated.

II. THEORY

II.a Assumptions and Approximations

The assumptions and approximations made in this investigation are described below.

1. The effective-mass approximation (EMA) was used for conduction band electrons in silicon, the oxide and all of the high-K dielectrics. However, it must be noted that the EMA can be used only when the structure is periodic.
2. The conservation of parallel momentum has been neglected in this work. This is based on experimental studies performed by Weinberg et. al. [8], which showed little difference between the gate currents on the (100) and (111) interfaces in silicon MOS structures.
3. A parabolic band structure was used in SiO_2 and the other high-K dielectrics. Because, the parabolic band structure does not give zero momentum at the dielectric valence band edge, in some of the past studies the E(k) dispersion proposed by Franz [9] has been used. However, the oxide effective mass remains a fitting parameter. Thus, a parabolic bandstructure has been used in this work due to lack of better knowledge of the bandstructure in the dielectric.
4. The impact of the image potential, as well as the actual barrier height of the dielectrics, are uncertain. Hence the barrier heights of the different dielectrics remain a fitting parameter.

II.b First-order perturbation method

In this section the self-consistent numerical solution of Poisson's and Schrodinger's equations is

0-7803-6279-9/00/$10.00 © 2000 IEEE

discussed. The tunneling current is calculated assuming that the wavefunction penetrates into the gate electrode on the left side and goes to zero at some point deep inside the substrate on the right side (Figure 1). Thus the solution of a quasi-bound system is required (A quasi-bound system is one in which the wavefunction does not go to zero at one of the boundaries). In order to overcome the open boundary condition at the oxide-gate electrode interface, the Quantum Transmitting Boundary Method is used.

In reality all Hamiltonians must be Hermitian. For a bound system the Hamiltonian can be written over the region in which it is bound. The matrix is Hermitian and the eigenvalues are real. However, if the system is open, it is not possible to write the Hamiltonian over the infinite region. In such cases, techniques such as QTBM [10] are used to write the Hamiltonian over a part of the domain. This leads to the introduction of complex elements in the Hamiltonian matrix, resulting in complex eigenenergy values. The physical meaning of the imaginary part of the eigenenergy can be

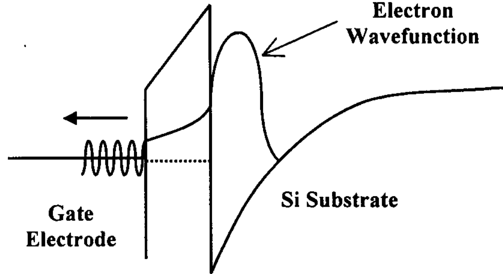

Figure 1: Energy band diagram, showing the electron tunneling into the gate electrode.

understood as the decay of the probability that the electron can be found inside the domain of description. Thus the real part E_R of the complex energy gives the energy of resonance, and its imaginary part Γ is related to the lifetime $\tau = \hbar/(2\Gamma)$ [11].

In this work Schrodinger's equation has been discretised using a finite difference technique. This discretisation leads to an infinite matrix since the system is quasi-bound. At the left boundary the general formalism of QTBM is used to connect the plane-wave solution for x<0 and the solution to the wave equation in the barrier region (Figure 1). This boundary condition introduces a complex number in the Hamiltonian matrix of the form A+iB. The real part, A, is of the same order of magnitude as the rest of the diagonal terms, while the imaginary part, B, is one order of magnitude lower and hence, can be treated as a perturbation. Thus the imaginary part of the eigenenergy is calculated as a perturbation.

For the initial caculations the Schrodinger's equation is solved self-consistently assuming that the wavefunction does not penetrate the gate dielectric. Subsequently, for the tunneling current calculations the Schrodinger's equation is *again* solved assuming that the wavefunction penetrates into the oxide. Thus, the effect of the wavefunction penetration into the oxide on the self-consistent electrostatic potential has been ignored. This assumption is valid as long as the contribution of the probability of the electron being inside the insulator is extremely small and is discussed in the Section III.b. This approximation facilitates the computing of the current in a post-processing fashion. The program computes the lifetime, $\Gamma_i = \mathrm{Im}(E_i)$, for the lowest three subbands as a post-processor. The product of the lifetime and the carrier concentration in each subband gives the tunneling current contribution from that subband. The total tunneling current is obtained using the following equation

$$J = \frac{q}{\hbar}\sum_i n_i \Gamma_i$$

where the sum is carried out over the lowest 3 subbands, and n_i is the electron sheet charge in the i-th subband. Fermi-Dirac statistics was used in computing n_i.

II.c Validity of the Perturbation Approach

In this section the validity of this approach to evaluate the direct tunneling current when the dielectric thickness becomes extremely small (0.5nm) is studied.

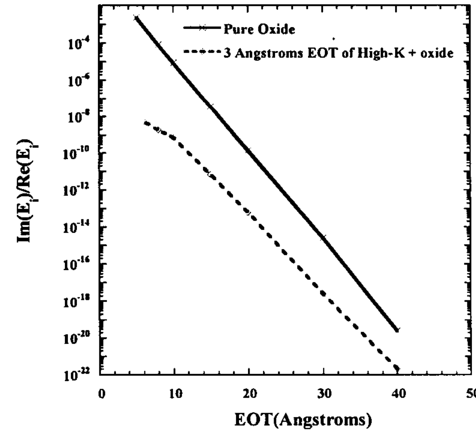

Figure 2: Plot of ratio of the imaginary part to the real part of the complex eigenenergy of the quasi-bound state for different gate dielectric stacks.

As described in the previous section, the imaginary part of the eigenenergy can be determined using a perturbation technique. This approach is valid as long as the magnitude of the imaginary part of the eigenenergy is much smaller than the magnitude of the real part. Hence simulations were performed to evaluate the imaginary part of the eigenenergy for a sample device biased at a gate voltage of 2V. A plot of the ratio of the imaginary part of the eigenenergy to the real part is shown in Figure 2, where it can be seen that the imaginary part is much smaller than the real part even for the worst case oxide thickness of 0.5nm. Figure 2 also

shows the same ratio for a high-K gate stack structure (1.7nm of oxide and 0.3nm EOT of a high-K dielectric with ε=23.4 corresponding to ZrO_2). The ratios are even smaller than that obtained for the pure oxide case. This is expected since the tunneling rate is proportional to the imaginary part, and it should decrease with the introduction of a high-K material. This verifies the validity of the perturbation approach.

III RESULTS AND DISCUSSIONS

In this section the simulated gate tunneling currents through different dielectric structures and the effects of wavefunction penetration into the gate stack on the quantization of carriers in the substrate is discussed.

III.a Tunneling Currents

As mentioned earlier, the band structure used for the different dielectrics is not realistic and hence the absolute value of the tunneling currents will not be accurate. However, the *relative* performance of different dielectrics can be studied and understood. The numerical solution technique discussed above has been implemented in UTQUANT, a self-consistent Schrodinger and Poisson solver [12].

A comparison of the tunneling currents through the high-K dielectric stack shown in Figure 2, calculated using the WKB method and the numerical method is shown in Figure 3.

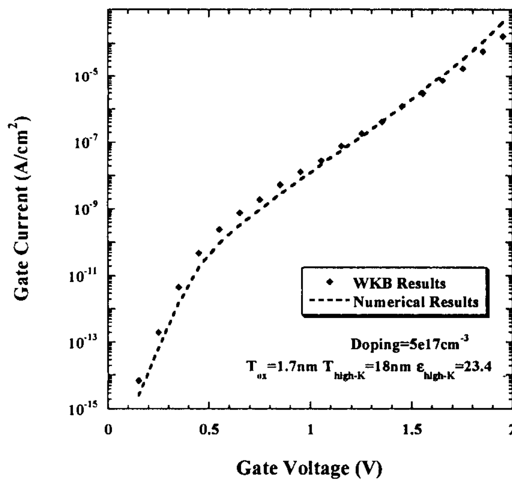

Figure 3: A comparison of the tunneling currents calculated using the WKB formalism and the numerical method.

Although the results compare reasonably well and the WKB technique seems to be applicable even in the high-K regime, the validity of WKB needs to be studied in more detail. A simulation of the currents through the various dielectrics was performed at 1V for an EOT of 1nm. The simulation could be performed only for those dielectrics for which the barrier height is known. A plot which summarizes the tunneling currents through different dielectrics is shown in Figure 4. It can be seen that the tunneling current decreases with increasing dielectric constant i.e., with increasing physical thickness. However, two-dimensional simulations have shown that the drain leakage current increases with increasing physical thickness [3, 4]. Hence the optimum dielectric must be chosen such that the gate leakage current as well as the drain leakage current is minimized.

In reality however, there is usually some residual interfacial oxide between the high-K dielectric and the silicon substrate. In order to study stacks of high-K dielectric and SiO_2, simulations were done with different gate stacks having 0.5nm of SiO_2 and 0.5nm EOT of different high-K dielectrics. It can be seen from Figure 4 that the tunneling currents through the stacks are much higher than that through the pure dielectric because the physical thickness is reduced. For example the physical thickness of a 1nm EOT of pure Ta_2O_5 dielectric is 6.4nm while the Ta_2O_5 stack has a physical thickness of only 3.7nm. Also, the tunneling

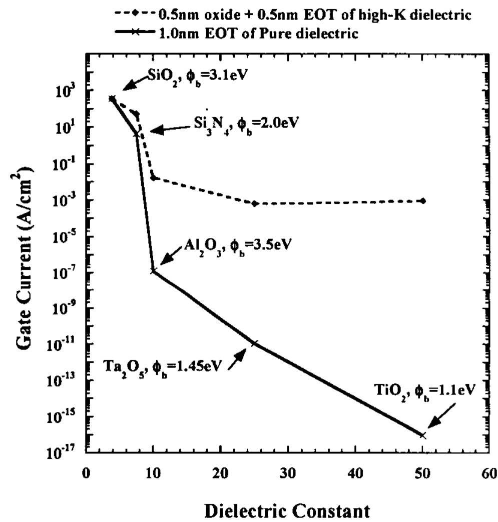

Figure 4: Simulated gate currents through different dielectrics of EOT 1.0nm at a gate voltage of 1V.

current is almost independent of the dielectric used for higher dielectric constants, because as the dielectric constant increases the increase in the physical thickness is compensated by a decrease in the barrier height. Therefore scaling of the dielectric is limited by the thickness of the interfacial oxide between the high-K and the silicon substrate.

III.b Effects of Wavefunction Penetration

Traditionally wavefunction penetration into the gate dielectric has been ignored in the study of quantization of carriers in the silicon substrate. Although this effect is

becoming more and more important with scaling of the gate oxide thickness to less than 2nm, this subject has hardly been discussed in the literature [5]. To better understand this, the Poisson solution was solved self-consistently with the Schrodinger's equation assuming that the wavefunction penetrates into the oxide and the gate electrode.

The results are very interesting. It is observed that the wave function shifts closer to the interface when the wavefunction is allowed to penetrate into the dielectric as shown in Figure 5. The capacitance thus calculated is higher than the capacitance derived assuming that the wavefunction does not penetrate the dielectric (Figure 6).

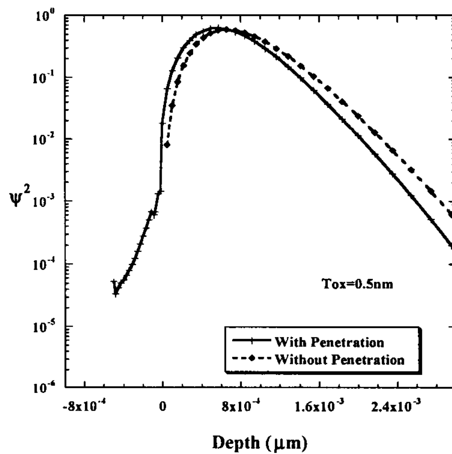

Figure 5.: Electron probability density calculated with and without the assumption that the wavefunction penetrates the gate dielectric.

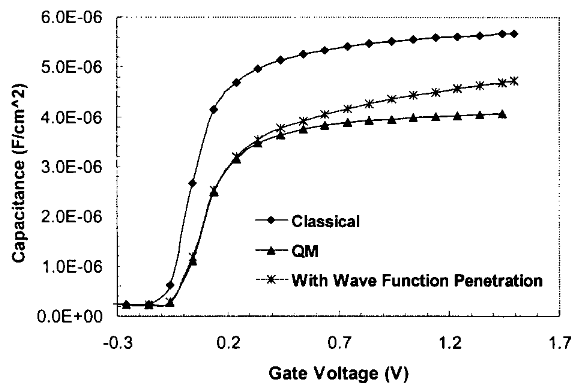

Figure 5.: The CV curves obtained with and without the assumption that the wavefunction penetrates the gate dielectric.

IV CONCLUSIONS

The direct tunneling currents through different high-K dielectrics have been investigated by numerically solving Schrodinger's equation, for the first time accounting for wavefunction penetration into the gate dielectric stack. The validity of the first-order perturbation approach has also been discussed. The tunneling currents predicted by this technique compare well with the WKB solution. However the validity of WKB in this regime needs to be studied in more detail. Furthermore in real devices, there is usually a thin film of residual oxide. In order to take full advantage of using high-K dielectrics as gate insulators the residual oxide needs to be eliminated. Finally, a study of the effects of wavefunction penetration on CV has been performed for the first time.

This work was supported in part by the Semiconductor Research Corporation FEP center.

REFERENCES

1. I. C. Kizilyalli et. al., "MOS Transistors with Stacked SiO_2-Ta_2O_5-SiO_2 Gate Dielectrics for Giga-Scale Integration of CMOS Technologies," IEEE Electron Device Letters, Vol. 19, No. 11, November 1998.
2. S. Krishnan et. al., "High-k Scaling of Gate Insulators: an Insightful Study," SPIE, p. 65, September 1998.
3. D. L. Kencke et. al., "Tinkering with the Well-Tempered MOSFET: Source-Channel Barrier Modulation with High-Permittivity Dielectrics," Vol. 27, p. 207, Superlattices and Microstructures.
4. E. M. Vogel et. al., "Modeled Tunnel Currents for High Dielectric Constant Dielectrics," IEEE Transactions on Electron Devices, Vol. 45, No. 6, June 1998.
5. F. Rana et. al., " Self-Consistent Modeling of Accumulation Layers and Tunneling Currents Through very Thin Oxides," Appl. Phys. Lett., vol.69, pp. 1104-1106, 1996.
6. W.-K. Shih et. al., "Modeling Gate Leakage Current in nMOS Structures Due to Tunneling Through an Ultra-Thin Oxide," Solid State Electronics, Vol. 42, No. 6, pp. 997-1006, 1998.
7. S.-H. Lo et. al., "Quantum Mechanical Modeling of Electron Tunneing from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's," IEEE Elec. Dev. Lett., No. 5, May 1997.
8. Z. A. Weinberg, "Tunneling of Electrons from Si into Thermally Grown SiO_2," Solid State Electronics, vol. 22, p.11, 1977.
9. M. Lenzlinger and E. H. Snow, "Fowler-Nordheim Tunneling into Thermally Grown SiO_2," J. Appl. Phys., vol. 40, p.278, 1969.
10. C. S. Lent and D. J. Kirkner, "The Quantum Transmitting Boundary Method," J. Appl. Phys. Vol. 67, p. 6353, May 1990.
11. T. B. Bahder et. al., "Resonant level lifetime in GaAs/AlGaAs double-barrier structures," Appl. Phys. Lett. 51, p. 1089, 1987.
12. W.-K. Shih, Ph.D. Dissertation, University of Texas at Austin, 1997.