Modeling of Direct Tunneling Current in Multi-layer Gate Stacks

Mohan V. Dunga, Xuemei Xi, Jin He, Igor Polishchuk, Qiang Lu, Mansun Chan, Ali M. Niknejad, and Chenming Hu

Department of EECS, University of California,Berkeley, CA 94720, USA

ABSTRACT

An analytical direct-tunneling gate current model for multi-layer gate dielectrics is presented. Theoretical derivation shows that the BSIM model for direct tunneling gate current through a single layer also works well for the multi-layer case. The theory is also supported by experimental data. This model is further extended to other modes of tunneling that occur at higher gate voltages. It has been shown that certain stack compositions result in higher leakage current depending on bias conditions. This model also predicts that the effectiveness of high-k dielectrics may decrease due to the reduction of band gap with increase in dielectric constant.

Keywords: gate current, direct tunneling, high-k dielectric, gate stack.

1 INTRODUCTION

Improvement in the performance of CMOS calls for aggressive device scaling with gate lengths down to sub-100nm regime. The resulting short-channel effects are controlled by a simultaneous reduction in the gate oxide thickness. The Semiconductor Industry Association (SIA) roadmap has projected the gate oxide thickness of the order of 2nm and below. As the gate oxide decreases below 2nm, the gate leakage increases due to direct tunneling leading to higher power consumption [1]-[2]. Since the leakage current increases exponentially with decrease in the physical gate oxide thickness, high-k dielectrics are being extensively studied in order to replace SiO_2 [3]-[4] which allow for higher physical thickness. In order to maintain good interface between the gate oxide and the channel, and to prevent mobility degradation, it is desirable to have a thin layer of SiO_2 between the bulk and the high-k dielectric [5]. Hence, the ITRS 2001 roadmap has called for the introduction of dielectric stacks and recognized the need for modeling to design the stacks [6]. Direct tunneling through single layer dielectric is well studied and an accurate analytical model for single layer direct tunneling exists [7]. Numerical methods also exist for tunneling through multi-layer stacks [8].

In this work an analytical model is developed for all regimes of gate tunneling through multi-layer stack. As the gate voltages increases, the tunneling mechanism changes

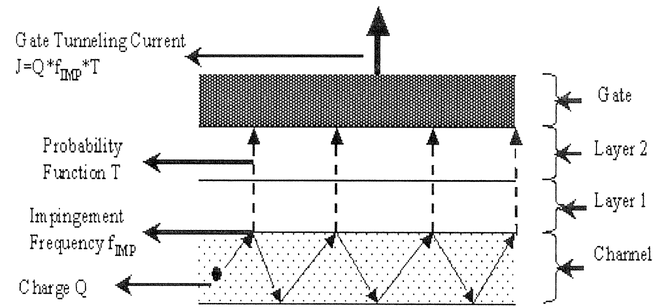


Figure 1: Components of tunneling current.

from direct tunneling to Fowler Nordheim (FN) tunneling. The model can be recast into single layer tunneling BSIM equation showing that the current BSIM model for single layer also works for the multi-layer case. The results show that introduction of high-k dielectric can lead to worse performance for certain Electrical Oxide Thickness (EOT) fractional compositions. As the dielectric constant of material increases, the band gap decreases. Thus, high-k dielectrics tend to have a lower barrier height. The model predicts that the performance of dielectric materials with high dielectric constants may be inferior unless they have high barriers to tunneling carriers.

2 THEORY

The model is based on the use of the well-known Wentzel-Kramers-Brillouin (WKB) approximation. The WKB approximation is used to calculate the tunneling probability of a carrier through an arbitrary barrier shape. The tunneling current (Fig. 1) is affected by the density of carriers (Q) available for injection from the inversion or the accumulation layer at the injection electrode, the rate at which these carriers impinge on the barrier (f_{imp}) and the probability rate (T) that these carriers tunnel through the barrier [9].

$$J = Q \times f_{imp} \times T \tag{1}$$

Q depends on the specific tunneling process [7]. It is a function of the gate voltage (V_g). The impingement frequency is a function of the vertical electric field, which is again a function of V_g. The tunneling probability is a function of both V_g and the barrier shape and composition.

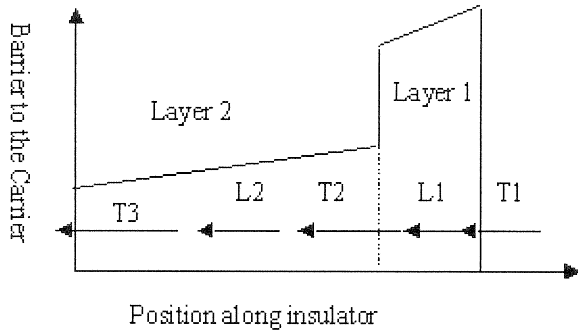


Figure 2: The probability function. Ti represents transmission factors at interfaces and Li represents loss factors in the layers.

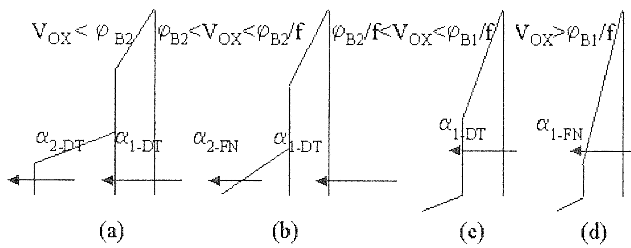


Fig. 3: Various tunneling processes. (a) DT-DT (b) DT-FN (c) DT (d) FN. The voltage regime and the calculation of T for each of the processes is shown. This enables calculation of J over wide range of V_{ox}.

For a single layer barrier, T is a product of the transmission factors at the interfaces and the loss factor L in the layer. The probability T is dominated by the loss factor which is a exponential function as given by the WKB approximation. This method is readily extended to the multi-layer case (Fig. 2) by including loss factors from other layers and transmission factors at new interfaces. The transmission factors are constant compared to loss factors. For the two layer case, T for direct tunneling is given by $Trans * \exp(-\alpha.toxe)$ where $Trans$ is the transmission factor and α is given by :

$$\frac{4}{3\hbar}\frac{toxe}{Vox}\left(\frac{K_1}{K_{ox}}\sqrt{2m_1 q}\left[\phi_{B1}^{3/2}-\left(\phi_{B1}-fV_{ox}\right)^{3/2}\right]+\frac{K_2}{K_{ox}}\sqrt{2m_2 q}\left[\left(\phi_{B2}-fV_{ox}\right)^{3/2}-\left(\phi_{B2}-V_{ox}\right)^{3/2}\right]\right) \tag{2}$$

where K, m and ϕ_B represent dielectric constant, effective mass of carrier and barrier to the carrier respectively for each layer and f is EOT fraction of the first layer. Direct tunneling occurs only when V_{ox} is smaller than ϕ_{B2}. In general, as the dielectric constant of the high-k materials increases, the band-gap decreases and the direct

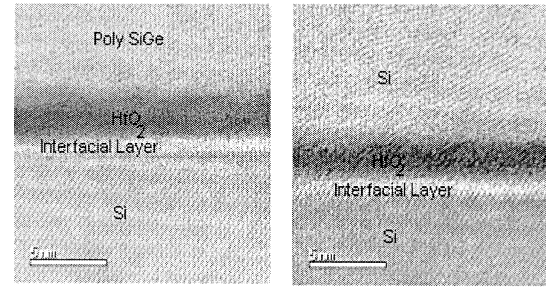
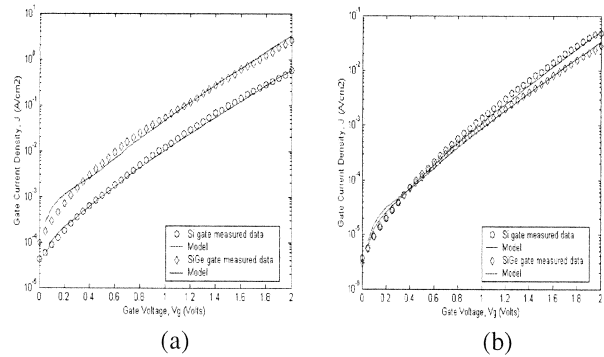


Fig 4: Cross-sectional TEM of HfO_2/Oxynitride gate stacks with Poly-SiGe gate and with Poly-Si gate.



(a) (b)

Fig. 5: The experimental and model gate current characteristics for gate dielectric stacks, (a) HfO_2/SiO_2 and (b) HfO_2/Oxynitride.

tunneling regime is reduced. At voltages higher than ϕ_{B2}, tunneling occurs by a two step process with direct tunneling through the first layer and FN tunneling through the second layer as shown in Fig. 3. Tunneling processes for even higher voltages and the corresponding voltage ranges are also shown. The WKB approximation is used to model the tunneling probability T in all the cases as shown in Fig. 3. The tunneling current J can be calculated in all cases by changing the tunneling probability function accordingly.

3 RESULTS AND DISCUSSIONS

To demonstrate the above model, the BSIM equation for direct tunneling through a single layer is used. The attenuation co-efficient α of the individual layers forming the stack can be expressed in terms of BSIM gate tunneling parameters. For the two-layer dielectric stack, we can obtain the new BSIM model parameters by expressing α_{new} (eqn. 2) in terms of α_1 and α_2 as :

$$\alpha_1 f+\alpha_2(1-f)+\frac{V_{ox}}{3\hbar}\left(K_1\sqrt{\frac{qm_1}{2\phi_{B1}}}-K2\sqrt{\frac{qm_2}{2\phi_{B2}}}\right)f(1-f) \tag{3}$$

Thus the single layer BSIM model extends to the two-layer case. By cascading layers, we can extend the BSIM model

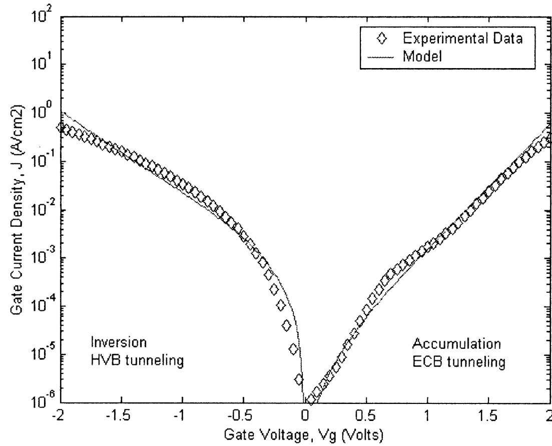


Fig. 6: The experimental and model gate current for the Si_3N_4/SiO_xN_y stack for PMOS devices.

Material	K	φ_B	Process
SiO_2	3.9	3.1	ECB
Si_3N_4	6.9	2.1	ECB
HfO_2	24	1.5	ECB
SiO_xN_Y	5.4	2.6	ECB
SiO_XN_Y	5.4	3.2	HVB
K_{new}	50	0.75	ECB

Table 1: Dielectric parameters used.

to the multi-layer stack. To verify the theory, the BSIM model was used to fit the experimental data. The gate currents of N-FETs with Poly-Si gate and with Poly-SiGe gate having HfO_2/SiO_2 stack and HfO_2/Oxynitride stack were measured under inversion. Poly-SiGe gate results in smaller EOT than poly-Si gate, but with surface nitridation the EOTs are closer. The cross-sectional transmission electron microscopy (TEM) gate-stack images (Fig. 4) of the devices with surface nitridation distinctly show the two layers forming the gate stack. Fig. 5 shows good agreement between the experimental data and the model for all devices. The tunneling process in this case was Electron tunneling from Conduction Band (ECB). Si_3N_4 is another commonly used high-k dielectric. Gate current was measured from P-FETs with $14A^0$ EOT fabricated using RTCVD. In the RTCVD process, a SiO_xN_y interfacial layer was grown followed by Si_3N_4 layer. Fig. 6 shows good agreement between the model and experimental data in both accumulation and inversion regimes. The tunneling process is Hole tunneling from Valence Band (HVB) in the inversion regime and is ECB in the accumulation regime. Table 1 shows the parameters associated with various dielectrics used in the calculations. The parameters for oxynitride are chosen using linear interpolation between parameters of SiO_2 and Si_3N_4. K_{new} is a imaginary high-k material exhibiting high-k and a low barrier height.

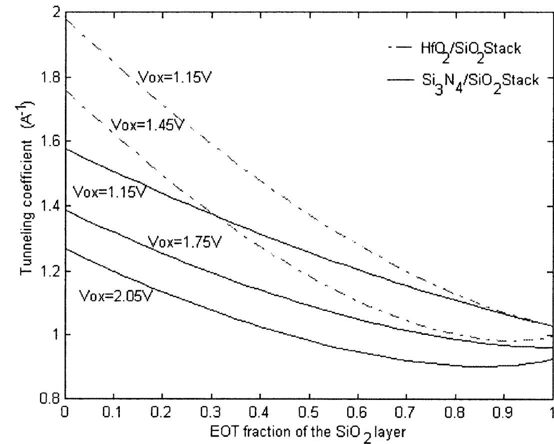


Fig. 7: Tunneling coefficient (A^{-1}) vs f for two different kind of stacks. For low V_{ox} there is no f_{crit}. For higher V_{ox}, f_{crit} follows eqn. (3).

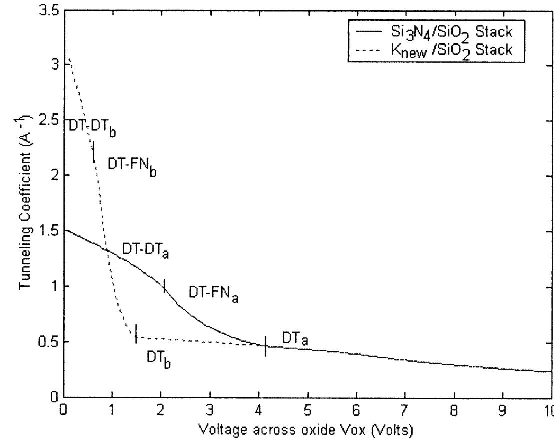


Fig. 8: Tunneling coefficient vs *Vox* for different kind of stacks. The tunneling regimes of the stacks indicate a smaller DT-DT window for K_{new}/SiO_2 stack.

The model was used to examine the variation of tunneling current as the EOT fraction of high-k dielectric is increased in a gate stack. In the direct tunneling regime, the tunneling probability does not decrease monotonically as expected. Fig. 7 shows that the tunneling probability increases with f and then decreases after some critical f (f_{crit}). The f_{crit} for a arbitrary stack is calculated from eqn. 2 and is given by :

$$f_{crit} = \frac{1}{V_{ox}}\left(\frac{\phi_{B1}K_1^2 m_1 - \phi_{B2}K_2^2 m_2}{K_1^2 m_1 - K_2^2 m_2}\right) \quad (4)$$

Eqn. 4 shows that for low operating voltages, f_{crit} can be greater than 1 implying a monotonic decrease in tunneling probability. Hence, the optimum fraction of high-k in a stack depends on the operating conditions. For a Si_3N_4/SiO_2

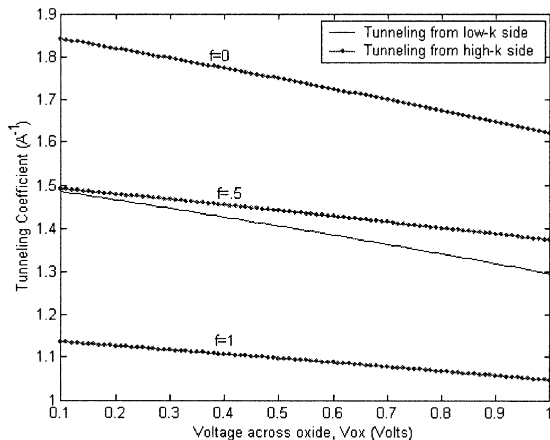


Fig: 9 Comparing the dependence of tunneling on the parameters of injection side. Si_3N_4/SiO_2 gate stack is used.

stack, f_{crit} is $1.756/V_{ox}$ and for an HfO_2/SiO_2 stack it is $1.33/V_{ox}$ (Fig. 7).

The model is also used to predict the tunneling probability beyond the direct tunneling regime. The tunneling probability is calculated as shown in Fig. 3. The tunneling coefficient is proportional to both K and φ_B. As K increases, generally φ_B decreases reducing the effectiveness of the high-k material. In addition, as seen from Fig. 3, the DT-DT regime is reduced as φ_{B2} decreases. Fig. 8 shows the tunneling coefficient for a Si_3N_4/SiO_2 gate stack with $f = .5$ over a large range of voltages. Tunneling coefficient for the imaginary high-k material K_{new} is also calculated. For low voltages, K_{new} is a better material but for large V_{ox} Si_3N_4 is a better dielectric. Thus there exists a trade-off between the dielectric constant and the barrier height, which will become more severe for materials with large dielectric constants.

Tunneling from either side of the barrier is compared in Fig. 9. When the gate is composed of a single layer, there is no difference between tunneling from either side. The coefficient is much larger for $f = 0$ compared to $f = 1$ resulting is much smaller leakage and illustrating the advantage of high-k materials. For an intermediate *f*, coefficients of tunneling from either side are unequal. The tunneling from high-k side (gate) is smaller than tunneling from the low-k side (substrate). This can also be inferred from eqn. 3, which shows that α is larger if layer 1 has a higher dielectric constant and a smaller barrier height compared to layer 2. As dielectric constant increases and barrier height decreases, the difference will increase and the tunneling of carriers from gate will be negligible.

4 CONCLUSION

A physical model for gate leakage current was developed. The model extends over a wide range of gate voltages covering both Direct Tunneling and Fowler Nordheim Tunneling regimes. The BSIM tunneling model is equivalent to this model. It was found that the model fits the experimental data well. The model showed that introduction of a certain fraction of high-k dielectric into the gate may lead to a worse performance at certain operating conditions. The trade-off between dielectric constant and barrier height was studied. It is necessary that future high-k dielectric materials should also have a relatively high barrier height to retain their efficacy. Tunneling from the gate terminal (high-k side) was found to be smaller compared to tunneling from substrate (high-k side). As the dielectric constant increases, this difference will become large and leakage restrictions will arise due to tunneling from the substrate side.

REFERENCES

[1] S.-H. Lo, D.Buchanan, Y. Taur, and W.Wang, "Quantum-Mechanical modelling of electron tunneling current from the inversion layer of ultra-thin-oxide MOSFET's," IEEE Electron Device Lett., vol. 18, p. 209, May 1997.

[2] F.Rana, S.Tiwari, and D.Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides," Appl. Phys. Lett., vol. 69, no. 8, pp. 1104-1106, 1996.

[3] T.P. Ma, "Making silicon nitride film a viable gate dielectric," IEEE Trans. Electron Devices, vol. 45, pp. 680-690, Mar. 1998.

[4] I.C.Kizilyalli, R.Y.S.Huang, and P.K.Roy, "MOS Transistors with Stacked SiO_2-Ta_2O_5-SiO_2 Gate Dielectrics for Giga-Scale Integration of CMOS Technologies," IEEE Electron Device Lett., vol. 19, no. 11, pp. 423-425, 1998.

[5] I. Polishchuk, and C. Hu, "Electron Wavefunction Penetration into Gate Dielectric and Interface Scattering - An Alternative to Surface Roughness Scattering Model," VLSI Tech. Symp., pp. 51-52, 2001.

[6] International Technology Roadmap for Semiconductors – Modeling and Simulation, Semiconduct. Ind. Assoc.,2001.

[7] W.en-Chin Lee and C. Hu, "Modeling CMOS Tunneling Currents Through Ultrathin Gate Oxide Due to Conduction- and Valence-Band Electron and Hole Tunneling," IEEE Trans. Elect. Dev, vol. 48, no. 7, pp. 1366-1373, July 2001.

[8] S. Mudanai, Y. Fan, Q. Ouyang, Al F. Tasch, and S. K. Banerjee, "Modeling of Direct Tunneling Current Through Gate Dielectric Stacks," IEEE Trans. Elect. Dev., vol. 47, no. 10, pp. 1851-1857, Oct. 2000.

[9] Igor Polishchuk, Y. Yeo, T. King, and C. Hu, "Tunneling through Multi-Layer Gate Dielectrics – An analaytical model," 60th DRC. Conference Digest, pp. 105-106, 2002.