

Tunneling currents through ultrathin oxide/nitride dual layer gate dielectrics for advanced microelectronic devices

H. Y. Yang, H. Niimi, and G. Lucovsky^{a)}

Departments of Electrical and Computer Engineering, Materials Science and Engineering, and Physics, North Carolina State University, Raleigh, North Carolina 27695-8202

(Received 8 May 1997; accepted for publication 3 November 1997)

Direct and Fowler–Nordheim tunneling currents through oxide and dual layer silicon oxide–silicon nitride dielectrics are investigated for substrate and gate injection. The calculations include depletion effects in the heavily doped (n^+) polysilicon gate electrodes as well as quantization effects in the less heavily doped n -type substrates. The Wentzel–Kramers–Brillouin (WKB) effective mass approximation has been compared with exact calculations for the tunneling probability, and based on these comparisons it has been found that the WKB approximation is adequate for single layer dielectrics, but is not for the dual layer dielectrics that are the focus of this article. Using exact tunneling transmission calculations, current-voltage (I – V) characteristics for ultrathin single layer oxides with different thicknesses (1.4, 2.0, and 2.3 nm) have been shown to agree well with recently reported experiments. Extensions of this approach demonstrate that direct tunneling currents in oxide/nitride structures with oxide equivalent thickness of 1.5 and 2.0 nm can be significantly lower than through single layer oxides of the same respective thickness. © 1998 American Institute of Physics. [S0021-8979(98)04304-7]

I. INTRODUCTION

In the last decade dynamic random access semiconductor memories have been the driving force for increasing the integration density in ultralarge scale integrated (ULSI) circuits, thereby moving silicon technology towards features sizes in the deep submicron regime (e.g., gate lengths $< 0.1 \mu\text{m}$ or 100 nm). As gate lengths are decreased to below 100 nm in advanced ULSI devices, gate dielectrics must be decreased to 3.0 nm or less.^{1,2} SiO_2 , the gate dielectric currently used in ULSI devices, shows several significant limitations in this thickness regime: (i) relatively high direct tunneling currents at low applied electrical fields ($\sim 1 \text{ A cm}^{-2}$), and/or (ii) boron diffusion into and through the oxide during dopant activation of p -type polysilicon gate electrodes in p -channel devices.^{3–5} In recent years stacked dielectrics, including oxide/nitride (ON) and oxide/nitride/oxide (ONO) composite structures have gained attention and have been proposed as an alternative to oxide gate dielectrics.^{6–10} The advantages of ON stacked structures over ultrathin SiO_2 films are superior reliability and prevention of boron diffusion. Consequently, a large amount of work has been done for films in the range of 5–10 nm.^{11–17} However, little is known about the current transport in stacked ON and ONO structures, when the oxide equivalent thicknesses of these dielectrics are reduced in the direct tunneling region extending to 1.5–2.0 nm.

A procedure for calculating the tunneling current for (ON) stacked dielectrics is developed, and subsequently applied to ultrathin films with oxide equivalent thickness of 1.5 and 2.0 nm. Polysilicon depletion in the gate electrodes and quantization effects in the substrate are treated exactly. Tun-

neling currents have been calculated as a function of (i) the oxide and nitride thicknesses, t_{ox} and t_{N} , respectively, (ii) the polysilicon doping density (N_{poly}), and (iii) the substrate doping density (N_{sub}). There are four materials-related parameters to be considered: the barrier heights at SiO_2 –Si and Si_3N_4 –Si interfaces, and the respective electron effective masses in the oxide and nitride layers. To determine the validity of the calculation and to obtain a value of the effective mass of tunneling electrons in SiO_2 , comparisons have been made between the calculations of this article, and experimental results for ultrathin single layer oxide films with thickness of 1.4, 2.0, and 2.3 nm as reported in Ref. 1. These calculations have used a value of 3.1 eV for the barrier height between the conduction bands of crystalline silicon and polysilicon, and SiO_2 .¹

II. ENERGY BAND DIAGRAMS

An energy band diagram for the device structure under consideration is shown in Fig. 1. It consists of an n^+ polysilicon gate electrode, a thin oxide/nitride dual layer gate dielectric, and a lightly doped n -type (n^-) crystalline silicon substrate. The oxide is in contact with the silicon substrate, and the nitride with the polysilicon gate electrode. For purposes of defining the relevant energy differences, the positions of the Fermi levels in the polysilicon and the silicon substrate have been taken to be the same, and the respective conduction bands are taken to be flat. Φ_{sub} is the energy difference between the oxide and substrate conduction bands, and Φ_{poly} is the energy difference between the nitride and polysilicon gate conduction bands. When a bias voltage (V_G) is applied to the polysilicon gate, the conduction and valence bands on both the polysilicon and substrate sides are bent. To present a complete description of the tunneling cur-

^{a)}Electronic mail: gerry_lucovsky@ncsu.edu

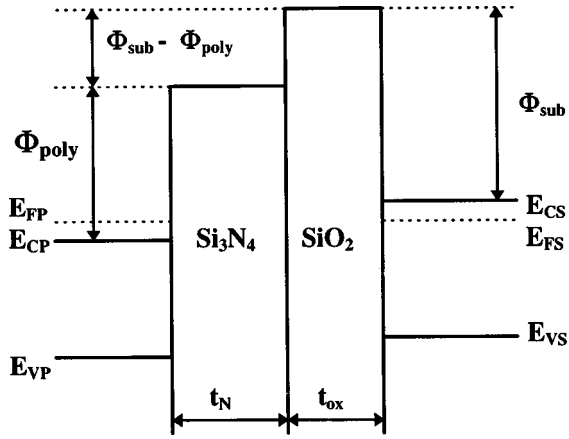


FIG. 1. Flat band diagram for $\text{SiO}_2/\text{Si}_3\text{N}_4$ dual dielectric with an n -type silicon substrate and a heavily doped n -type polysilicon gate electrode. The diagram emphasizes the respective conduction band barrier heights, Φ_{poly} and Φ_{sub} , and the potential step internal to the dual layer dielectric film, $\Phi_{\text{sub}} - \Phi_{\text{poly}}$. The thicknesses of the oxide and nitride are t_{ox} and t_{N} , respectively.

rents, which must include band bending effects in the gate electrode, it is necessary to consider differences between bias voltages which are associated with substrate injection (V_G positive) and gate injection (V_G negative) of tunneling electrons.

For substrate injection of electrons, a positive voltage is applied to polysilicon gate, and the conduction band in the substrate bends downward causing a strong accumulation of electrons near the semiconductor surface. In contrast, the conduction band in the polysilicon gate bends upward resulting in carrier depletion near the polysilicon surface. For gate injection, the situation is reversed. The extent of the band bending in the polysilicon and substrate increase as the applied voltage increases. This is quantified by partitioning the applied potential across (i) the polysilicon gate, (ii) the oxide/nitride stacked dielectric, and (iii) the substrate in serial manner; i.e.,

$$V_G = V_{\text{FB}} + V_{\text{ox}} + V_{\text{N}} + \psi_{\text{sub}} + \psi_{\text{poly}}, \quad (1)$$

where V_{FB} is flat band voltage; i.e., the difference in the Fermi level positions between the polysilicon and silicon substrate, V_{ox} and V_{N} are the respective potential drops in the oxide and nitride layers, and ψ_{sub} and ψ_{poly} are the respective potential drops in the substrate and polysilicon.

For stacked oxide/nitride gate dielectrics, the tunneling mechanism can be either direct (DT) or Fowler–Nordheim (FNT), or a combination of both depending on the magnitude and sign of the applied voltage. The distinction between FN tunneling and direct tunneling of electrons is defined by the shape of the tunneling barrier. If the oxide or nitride tunneling barrier is triangular and the transported electrons transit in part through condition band states, then FN tunneling is said to occur. If the barrier is trapezoidal and the electrons do not transit through conduction band states, then the tunneling is said to be direct. For substrate injection, there are four distinct tunneling regimes defined by the applied voltages shown in Fig. 2:

- (i) Fig. 2(a): Direct tunneling in both oxide and nitride films,

$$\Phi_{\text{sub}} + \psi_{\text{sub}} - E \geq V_{\text{ox}} + V_{\text{N}} + \Phi_{\text{sub}} - \Phi_{\text{poly}}; \quad (2)$$
- (ii) Fig. 2(b): Direct tunneling in the oxide and FN in nitride,

$$V_{\text{ox}} + \Phi_{\text{sub}} - \Phi_{\text{poly}} < \Phi_{\text{sub}} + \psi_{\text{sub}} - E < V_{\text{ox}} + V_{\text{N}} + \Phi_{\text{sub}} - \Phi_{\text{poly}}; \quad (3)$$
- (iii) Fig. 2(c): Direct tunneling in the oxide layer,

$$V_{\text{ox}} < \Phi_{\text{sub}} + \psi_{\text{sub}} - E < V_{\text{ox}} + \Phi_{\text{sub}} - \Phi_{\text{poly}}; \quad \text{and} \quad (4)$$
- (iv) Fig. 2(d): FN tunneling in oxide layer,

$$\Phi_{\text{sub}} + \psi_{\text{sub}} - E < V_{\text{ox}}, \quad (5)$$

where E is the total energy of the tunneling electron measured from the substrate conduction band (CB) edge at the Si/SiO_2 interface. The four tunneling conditions presented above are for monotonically increasing applied gate voltages. The changes in tunneling represented by Eqs. (2)–(5) are independent of ratio of oxide to nitride layer thickness.

For gate injection, there are also four different tunneling conditions as shown in Fig. 3; however unlike the regimes defined by Eqs. (2)–(5), these do not represent a sequence of increasing applied bias that is independent of the oxide to nitride thickness ratio. The situation presented in Eq. (6) is for the lowest applied bias voltage, and the situations presented in Eqs. (7)–(9) are for higher applied bias voltages, but are not part of a monotonically increasing sequence of applied biases. Such a sequence must be defined by a particular ratio of oxide to nitride thicknesses, and in general will include at least one of the conditions in Eqs. (7), (8), or (9). The four different conditions shown in Figs. 3(a)–3(d) are given by

- (i) Fig. 3(a): Direct tunneling in the oxide and nitride layers,

$$\Phi_{\text{poly}} + \psi_{\text{poly}} - E \geq \max(V_{\text{N}}, V_{\text{ox}} + V_{\text{N}} - \Phi_{\text{sub}} + \Phi_{\text{poly}}); \quad (6)$$
- (ii) Fig. 3(b): Direct tunneling in the oxide and FN in nitride,

$$V_{\text{N}} < \Phi_{\text{poly}} + \psi_{\text{poly}} - E < V_{\text{ox}} + V_{\text{N}} - \Phi_{\text{sub}} + \Phi_{\text{poly}}; \quad (7)$$
- (iii) Fig. 3(c): FN in oxide and direct tunneling in nitride,

$$V_{\text{ox}} + V_{\text{N}} - \Phi_{\text{sub}} + \Phi_{\text{poly}} < \Phi_{\text{poly}} + \psi_{\text{poly}} - E < V_{\text{N}}; \quad \text{and} \quad (8)$$
- (iv) Fig. 3(d): FN tunneling in both the oxide and nitride,

$$\Phi_{\text{poly}} + \psi_{\text{poly}} - E < \min(V_{\text{N}}, V_{\text{ox}} + V_{\text{N}} - \Phi_{\text{sub}} + \Phi_{\text{poly}}), \quad (9)$$

where E again is the total energy of the tunneling electron measured from the polysilicon CB edge at the polysilicon/ Si_3N_4 interface.

III. CALCULATION OF TUNNELING CURRENTS

This section discusses the tunneling current calculations. First, the tunneling current density J is introduced; second, the band bending with quantization effects for both substrate accumulation and inversion are presented; third, the necessity of including polysilicon depletion is demonstrated;

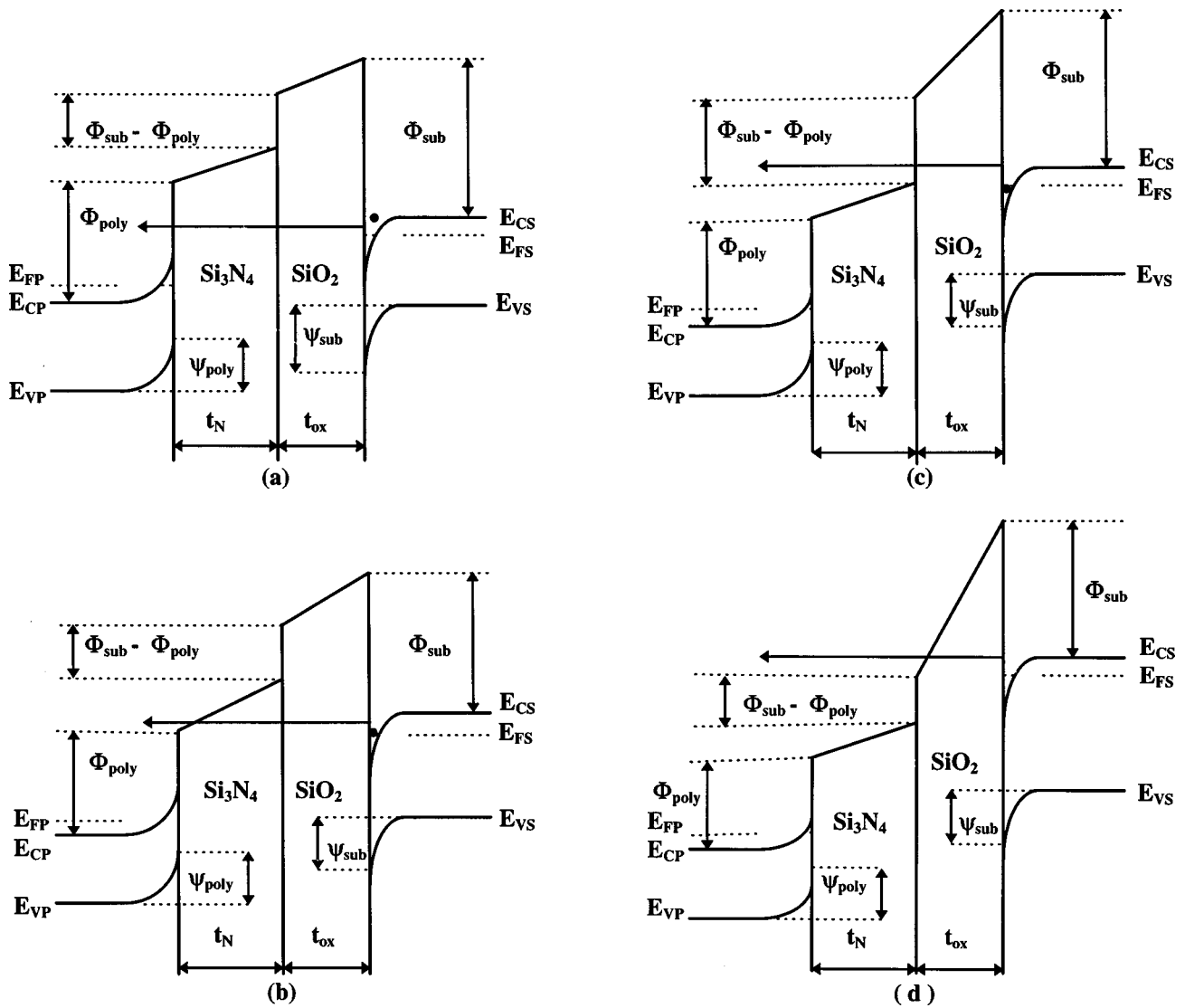


FIG. 2. Band configuration for substrate injection for a dual layer SiO₂ and Si₃N₄ dielectric. The applied bias is the difference in the Fermi level positions, $E_{FS} - E_{FP}$, and Ψ_{sub} and Ψ_{poly} are the voltage drops in the substrate and polysilicon, respectively. The applied bias increases in going from (a) to (d). (a) Band configuration at a low applied bias such that direct tunneling takes place through both the SiO₂ and Si₃N₄ layers. (b) Band configuration for a higher applied bias such that direct tunneling occurs through the SiO₂ and Fowler–Nordheim tunneling occurs through the Si₃N₄. (c) Band configuration at a still higher bias so that direct tunneling takes place only in through the SiO₂ film. (d) Band configuration at the highest applied bias of this sequence so that Fowler–Nordheim tunneling takes place only in through the SiO₂ film.

fourth, the relationships to calculate the flat band voltage and Fermi levels are described; and finally, the relationship between V_{ox} and V_N is given.

A. Tunneling current density: J

The tunneling current density for substrate injection and gate injection are calculated using the independent electron approximation and assuming an elastic tunneling process.¹⁸ It is assumed that the transverse component of the electron energy (E_t) is conserved during tunneling through the oxide/nitride dual layer of the device structure. If a parabolic dispersion relation is used for the transverse energy component with a transverse electron effective mass (m_t) the tunneling current (J) is then given by^{19,20}

$$J = \frac{4\pi q m_t}{h^3} \int_0^{E_{ES}} \left[\int_0^E T_t(E, E_t) dE_t \right] dE, \tag{10}$$

where E is the total energy of the tunneling electron measured from the substrate CB edge at the Si/SiO₂ interface, and T_t the tunneling transmission probability. The quantum mechanical probability for oxide/nitride dual layer structure can be calculated using the Wentzel–Kramers–Brillouin (WKB) approximation,²¹

$$T_t(E, k_t) = \exp \left\{ -\frac{4\pi}{h} \left[\int_0^{x_1} k_{ox}^{im}(E, k_t, x) + \int_{x_1}^{x_2} k_N^{im}(E, k_t, x) \right] dx \right\}, \tag{11}$$

where k_{ox}^{im} and k_N^{im} are the imaginary parts of the complex electron wave vector of the tunneling electrons within the oxide and nitride, respectively, x_1 is the distance from Si/SiO₂ interface to the classical turning point of the oxide, and x_2 is the distance from x_1 to the classical turning point of

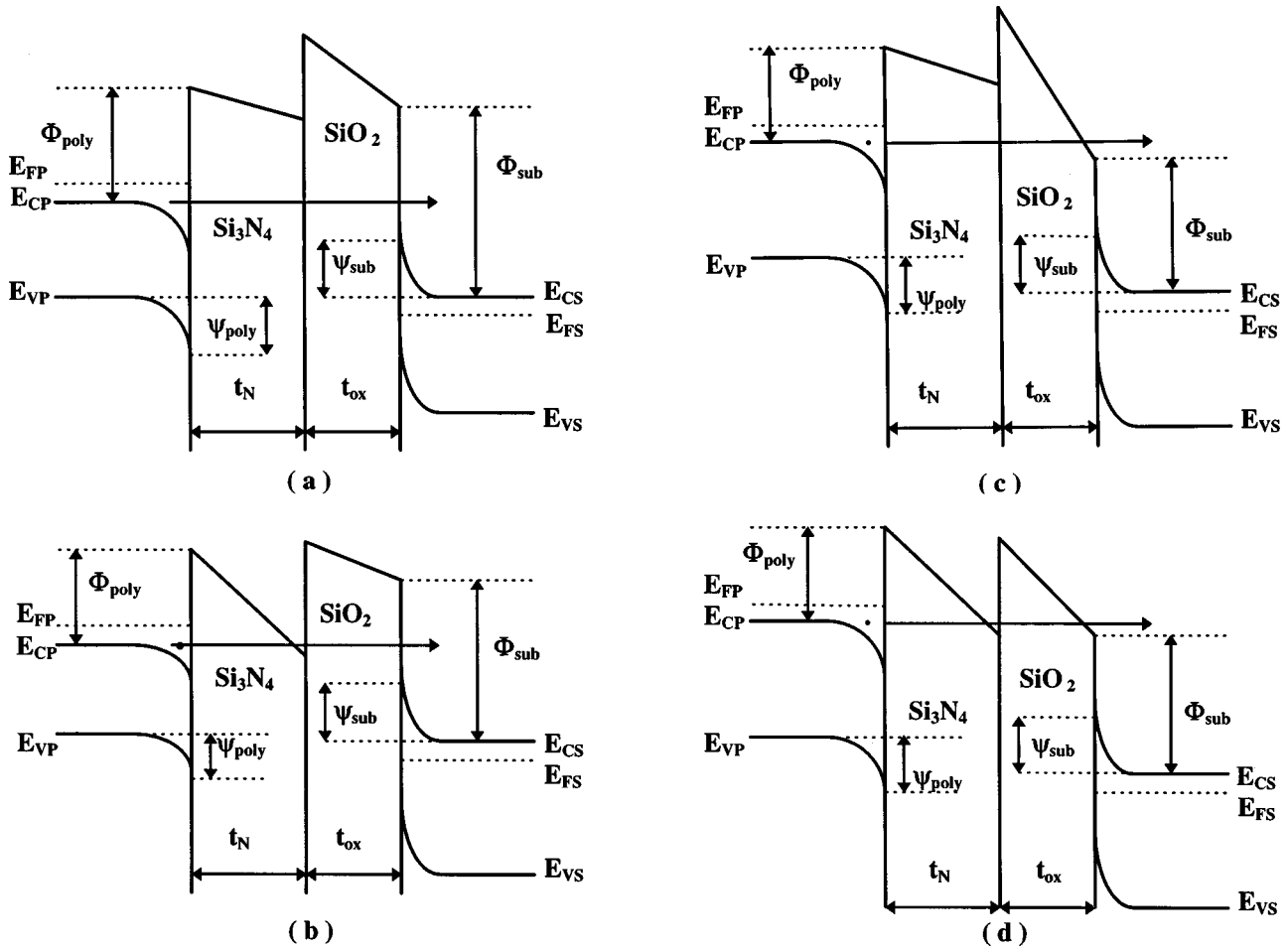


FIG. 3. Band configuration for gate injection for a dual layer SiO₂ and Si₃N₄ dielectric. The applied bias is the difference in the Fermi level positions, $E_{FP} - E_{FS}$, and Ψ_{sub} and Ψ_{poly} are the voltage drops in the substrate and polysilicon, respectively. (a) Band configuration at a low applied bias such that direct tunneling takes place through both the SiO₂ and Si₃N₄ layers, and as noted in the text do not represent a monotonically increasing sequence of voltages, but do represent higher applied bias levels than shown in (a). (b) Band configuration for an applied bias at which Fowler–Nordheim tunneling occurs through the Si₃N₄ and direct tunneling occurs through the SiO₂. (c) Band configuration at an applied bias at which direct tunneling takes place in the Si₃N₄ and Fowler–Nordheim tunneling in the SiO₂ film. (d) Band configuration at an applied bias at which Fowler–Nordheim tunneling takes through both the Si₃N₄ and the SiO₂ films.

nitride. For direct tunneling, x_1 is equal to the oxide thickness (t_{ox}) while for FN tunneling, x_1 is determined by the turning point at the SiO₂ conduction band edge. Similarly, for directing tunneling x_2 is equal to oxide and nitride thickness, $t_{ox} + t_N$, while for FN tunneling, x_2 is determined by the turning point at the nitride conduction band edge.

The values for k_{ox}^{im} and k_N^{im} can be calculated by considering the wave vectors, k_{ox} and k_N , of the tunneling electron as the total of the constant transverse component, k_t , and the normal imaginary parts, k_{ox}^{im} and k_N^{im} , respectively, that is:

$$k_{ox} = k_t + ik_{ox}^{im} \tag{12}$$

$$k_N = k_t + ik_N^{im}.$$

A one-band parabolic dispersion relationship was used to calculate k_{ox} and k_N :

$$k_{ox}^2(x) = \frac{8\pi^2 m_{ox}}{h^2} [E - E_{ox}^C(x)], \tag{13}$$

$$k_N^2(x) = \frac{8\pi^2 m_N}{h^2} [E - E_N^C(x)],$$

where $E_{ox}^C(x)$ and $E_N^C(x)$ are the energies of the oxide and nitride conduction band edges, respectively, at a distance x from the Si/SiO₂ interface, m_{ox} and m_N are the effective mass of the tunneling electron in the oxide and nitride, respectively. From Eqs. (11)–(13), the tunneling transmission, T_t , can be expressed as a function of the total energy, E , the transverse component, E_t as well as $E_{ox}^C(x)$ and $E_N^C(x)$:

$$T_1(E, E_t) = \exp \left\{ -\frac{4\pi}{h} \left[\int_0^{x_1} \sqrt{2m_t E_t - 2m_{ox} [E - E_{ox}^C(x)]} \right. \right. \\ \left. \left. + \int_{x_1}^{x_2} \sqrt{2m_t E_t - m_N [E - E_N^C(x)]} \right] dx \right\}. \tag{14}$$

From this expression and Eq. (10), the tunneling current density J can then be calculated.

For substrate injection, the energies $E_{ox}^C(x)$ and $E_N^C(x)$ of the oxide and nitride conduction bands, respectively, are given by

$$E_{\text{ox}}^C(x) = \Phi_{\text{sub}} + \Psi_{\text{sub}} - E - q \frac{V_{\text{ox}}}{t_{\text{ox}}} x,$$

$$E_{\text{N}}^C(x) = \Phi_{\text{sub}} + \Psi_{\text{sub}} - E - q V_{\text{ox}} - (\Phi_{\text{sub}} - \Phi_{\text{poly}}) - q \frac{V_{\text{N}}}{t_{\text{N}}} (x - t_{\text{ox}}). \tag{15}$$

For the four cases of substrate injection, x_1 and x_2 for both oxide and nitride DT or FNT, or a combination of both are determined as follows:

(i) Fig. 2(a): DT(in SiO₂): $x_1 = t_{\text{ox}}$
 DT(in Si₃N₄): $x_2 = t_{\text{ox}} + t_{\text{N}}$, (16)

(ii) Fig. 2(b): DT(in SiO₂): $x_1 = t_{\text{ox}}$
 FNT(in Si₃N₄): $x_2 = t_{\text{ox}} + \frac{\Phi_{\text{sub}} + \Psi_{\text{sub}} - q V_{\text{ox}} - (\Phi_{\text{sub}} - \Phi_{\text{poly}}) - E}{q V_{\text{N}}} t_{\text{N}}$, (17)

(iii) Fig. 2(c): DT(only in SiO₂): $x_1 = x_2 = t_{\text{ox}}$, and (18)

(iv) Fig. 2(d): FNT(only in SiO₂): $x_1 = x_2 = \frac{\Phi_{\text{sub}} + \Psi_{\text{sub}} - E}{q V_{\text{ox}}} t_{\text{ox}}$. (19)

For gate injection, Eq. (15) is changed to

$$T_1(E, E_t) = \exp \left\{ -\frac{4\pi}{h} \left[\int_0^{x_1} \sqrt{2m_t E_t - 2m_{\text{N}} [E - E_{\text{N}}^C(x)]} + \int_{x_1}^{x_2} \sqrt{2m_t E_t - 2m_{\text{ox}} [E - E_{\text{ox}}^C(x)]} dx \right] \right\}, \tag{20}$$

and the energies $E_{\text{ox}}^C(x)$ and $E_{\text{N}}^C(x)$ of the oxide conduction band and the nitride conduction bands, respectively, at a distance x from poly-Si/Si₃N₄ interface are given by

$$E_{\text{N}}^C(x) = \Phi_{\text{poly}} + \Psi_{\text{poly}} - E - q \frac{V_{\text{N}}}{t_{\text{N}}} x,$$

$$E_{\text{ox}}^C(x) = \Phi_{\text{poly}} + \Psi_{\text{poly}} - E - q V_{\text{N}} + (\Phi_{\text{sub}} - \Phi_{\text{poly}}) - q \frac{V_{\text{ox}}}{t_{\text{ox}}} (x - t_{\text{N}}). \tag{21}$$

Similarly, for gate injection, x_1 and x_2 for both oxide and nitride DT or FNT, or a combination of both are determined as follows:

(i) Fig. 3(a): DT(in Si₃N₄): $x_1 = t_{\text{N}}$
 DT(in SiO₂): $x_2 = t_{\text{ox}} + t_{\text{N}}$, (22)

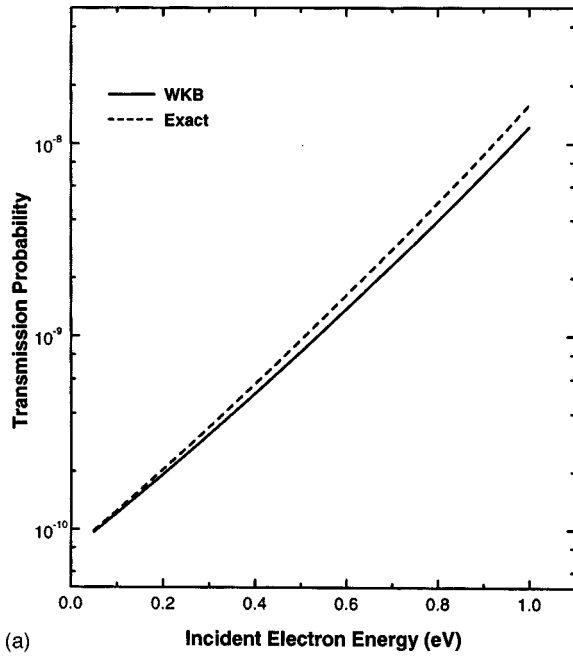
(ii) Fig. 3(b): DT(in Si₃N₄): $x_1 = t_{\text{N}}$
 FNT(in SiO₂): $x_2 = t_{\text{N}} + \frac{\Phi_{\text{poly}} + \Psi_{\text{poly}} - E - [q V_{\text{ox}} - (\Phi_{\text{sub}} - \Phi_{\text{poly}})]}{q V_{\text{ox}}} t_{\text{ox}}$, (23)

(iii) Fig. 3(c): FNT(in Si₃N₄): $x_1 = t_{\text{N}} + \frac{\Phi_{\text{poly}} + \Psi_{\text{poly}} - E}{q V_{\text{N}}} t_{\text{N}}$
 DT(in SiO₂): $x_2 = t_{\text{N}} + t_{\text{ox}}$, (24)

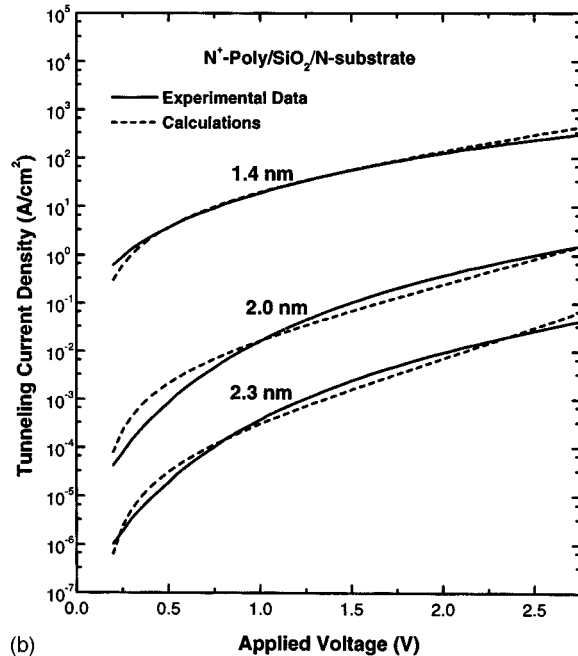
(iv) Fig. 3(d): DT(in Si₃N₄): $x_1 = \frac{\Phi_{\text{poly}} + \Psi_{\text{poly}} - E}{q V_{\text{N}}} t_{\text{N}}$
 FNT(in SiO₂): $x_2 = t_{\text{N}} + \frac{\Phi_{\text{poly}} + \Psi_{\text{poly}} - E - [q V_{\text{ox}} - (\Phi_{\text{sub}} - \Phi_{\text{poly}})]}{q V_{\text{ox}}} t_{\text{ox}}$. (25)

Figures 4(a) and 5(a) compare, respectively, the transmission probabilities for direct tunneling as a function of the incident energy for single layer oxide and dual layer oxide–nitride dielectrics with oxide equivalent thicknesses equal to 2.0 nm. These comparisons demonstrate that the WKB approximation is adequate for single layer dielectrics, but that

it underestimates the transmission probability for the dual layer structures. Based on these comparisons, we have elected to use the exact transmission probabilities in the calculations that follow. This means that the exact transmission probabilities, rather than the WKB approximations must be used in Eq. (10).



(a)



(b)

FIG. 4. (a) Comparisons between exact and WKB tunneling transmission coefficients for single layer 2.0 nm oxides. (b) Comparison between substrate injection experimental and calculated tunneling currents through single SiO₂ layer gate dielectrics. The *n*-type substrate doping is $1.0 \times 10^{17} \text{ cm}^{-3}$, and the *n*-type polycrystalline silicon gate doping is $5.0 \times 10^{19} \text{ cm}^{-3}$. The dielectric constant of the oxide layer is 3.8. The offset between the conduction band of each silicon layer and the conduction band of the SiO₂ film is 3.1 eV, and the effective mass for electrons in the SiO₂ film is $0.5 m_e$. The values of silicon substrate-oxide band offset and oxide effective mass are also used in Figs. 5, 6, 7, and 11.

B. Quantization effects in the substrate

In order to model the tunneling current for the stacked oxide/nitride dual layer gate dielectrics, it is necessary to consider quantization effects within the accumulated and inverted layers of the substrate. Extensive studies of these ef-

fects have been made,^{22–29} and the equations for calculations of band bending, Ψ_s , used in this article are given by Refs. 24 and 26:

(i) For inversion:

$$\Psi_s = \frac{55}{96} \left(\frac{32}{11} \right)^{1/3} \left(\frac{3}{2} \right)^{5/3} \left(\frac{q\hbar V_{\text{ox}}}{m_{\text{Si}}^{1/2} t_{\text{ox}}} \right)^{2/3}, \quad (26)$$

(ii) For accumulation:

$$\Psi_s = \left(\frac{9\hbar^2}{8m_{\text{Si}}} \right)^{1/3} \left(\frac{1.754\pi q \epsilon_{\text{ox}} V_{\text{ox}}}{\epsilon_{\text{Si}} t_{\text{ox}}} \right)^{2/3}, \quad (27)$$

where m_{Si} is the effective mass in the silicon substrate for motion perpendicular to the interface.²⁵

C. Polysilicon depletion effects

In addition to the voltage drop in the accumulated and inverted substrate layers, the voltage drop in the depleted or inverted polysilicon gate has to be taken into account. As far as polysilicon band bending, ψ_{poly} , is concerned, in depletion the effects of quantization can be ignored because the concentration of electrons and holes is negligible in comparison with that of ionized donors. Hence, the classical formula^{30–32}

$$\Psi_{\text{poly}} = \frac{\epsilon_{\text{N}}^2 V_{\text{N}}^2}{2q \epsilon_{\text{poly}} N_{\text{poly}} t_{\text{N}}^2} \quad (28)$$

is valid. When the polysilicon gate is inverted, the quantization of the motion of holes should be in principle taken into account. Nevertheless, the charge associated with these holes is exponentially dependent on ψ_{poly} in strong inversion. This voltage drop remains practically constant, and it is assumed that $\psi_{\text{poly}} = 1.12 \text{ eV}$, the band gap of silicon. Thus we will assume that ψ_{poly} is given by Eq. (27) for $\psi_{\text{poly}} < 1.12 \text{ eV}$, whereas $\psi_{\text{poly}} = 1.12 \text{ eV}$ for larger applied bias voltages. This approximation is usually made in the classical approach, and remains acceptable in the quantum mechanical model as well.

D. Flatband voltage and Fermi level position

If interface and oxide charges are neglected, the flat band voltage drop, V_{FB} , between the polysilicon and the substrate is equal to their work function difference as given by

$$V_{\text{FB}} = \frac{k_{\text{B}} T}{q} \log \left(\frac{N_{\text{sub}}}{N_{\text{poly}}} \right). \quad (29)$$

The energy difference between the bottom of the conduction band E_c and the Fermi level (E_F) for the substrate is calculated by using the Fermi integral of order 1/2 (Ref. 33)

$$N_{\text{sub}} = N_c F_{1/2}(\Delta), \quad (30)$$

where

$$N_c = 2 \left(\frac{m_{\text{dn}} k_{\text{B}} T}{2\pi\hbar^2} \right)^{3/2}, \quad (31)$$

$$\Delta = \left(\frac{E_F - E_c}{k_{\text{B}} T} \right), \quad \text{and} \quad (32)$$

$$F_{1/2}(\Delta) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2} dx}{[1 + \exp(x - \Delta)]}. \quad (33)$$

For heavily doped polysilicon, the Fermi level lies within the conduction band and polysilicon gate electrode is a degenerate semiconductor. In this case³⁴

$$E_F - E_C = \frac{(3\pi^2 N_{\text{poly}})^{2/3} \hbar^2}{2m_{\text{dn}}}, \quad (34)$$

where m_{dn} is the density of states effective mass.

E. The relationship between V_{ox} and V_{N}

The relationship between V_{ox} and V_{N} is simply determined from the application of Gauss's law

$$\frac{\epsilon_{\text{N}} V_{\text{N}}}{t_{\text{N}}} = \frac{\epsilon_{\text{ox}} V_{\text{ox}}}{t_{\text{ox}}}. \quad (35)$$

F. Image force lowering of interfacial potential barriers

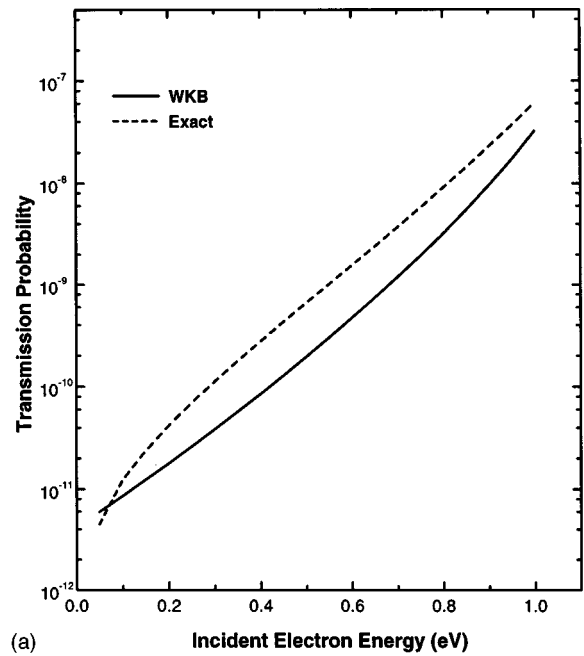
The effects of image force lowering at the Si-dielectric interfaces have been investigated in the classical approximation, and have been found to be negligible supporting the approaches used in other treatments of tunneling through single layer oxides.³⁵

IV. RESULTS AND DISCUSSION

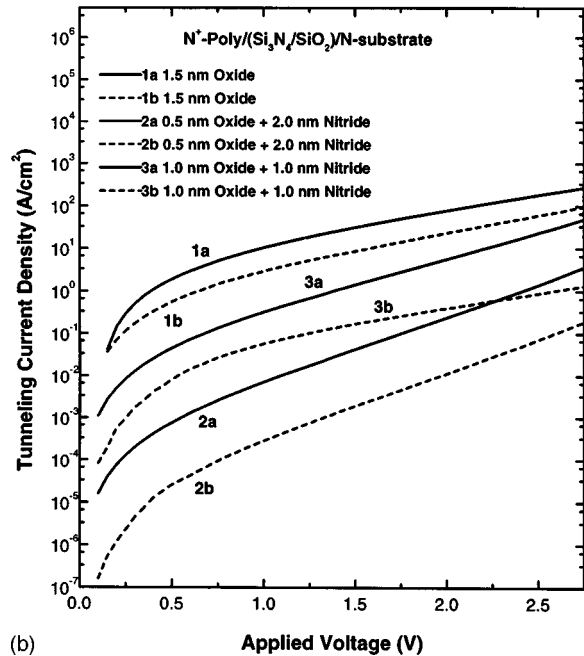
In this section we apply the formalism developed above to calculate the tunneling currents for different ultrathin gate electrodes. For the purposes of these calculations it is initially assumed that the energy dispersion relation in the mid-gap region of SiO_2 and Si_3N_4 is described by the same value of the effective mass $m_{\text{ox}} = m_{\text{N}}$. The barrier heights due to conduction band discontinuity at the Si/ SiO_2 interface and the poly-Si/ Si_3N_4 interface are 3.1 and 2.1 eV, respectively. As noted above the effects due to image force barrier lowering are negligibly small in the thickness range of interest and therefore have not been incorporated into the calculations.³⁵

Figure 4(b) shows the calculated substrate injection tunneling current versus applied gate voltage at 300 K for single layer oxides of thicknesses 1.4, 2.0, and 2.3 nm. This calculation is made by setting $t_{\text{N}} = 0$ and $\Phi_{\text{sub}} = \Phi_{\text{poly}} = 3.1$ eV. To fit the experimental results of Refs. 1 and 36, m_{ox} has been set equal to $0.5m_e$. For the voltage range shown, conduction occurs in the direct tunneling regime. The same value of m_{ox} gives good agreement with the experimental results for all three oxide thicknesses, thereby spanning a current density range of nine orders of magnitude from approximately 10^{-6} A/cm² to 10^3 A/cm²; this corresponds to a voltage range of 2.75 V. The values of doping used for the substrate and gate electrode are those identified in Refs. 1 and 36.

Figures 5(b) and 6 present the results of similar calculations for both substrate and gate injection tunneling currents for oxide/nitride dual layer gate dielectrics with oxide equivalent thickness of 1.5 [Fig. 5(b)] and 2.0 nm (Fig. 6). For purposes of comparison, the respective figures also include tunneling current calculations for single layer oxides of thickness 1.5 and 2.0 nm. As noted above the calculations for



(a)



(b)

FIG. 5. (a) Comparisons between exact and WKB tunneling transmission coefficients for dual layer oxide-nitride dielectrics with an equivalent oxide thickness of 2.0 nm, and equal oxide and nitride layer thicknesses, ~ 1.3 nm. (b) Comparison between calculated tunneling currents for substrate and gate injection tunneling currents through oxide/nitride double layer gate dielectrics with an equivalent oxide thickness of 1.5 nm with a single oxide layer of thickness 1.5 nm. Solid lines are for substrate injection, and dashed lines for gate injection. The dielectric constants are 3.8 for the oxide and 7.6 for the nitride, respectively. The polysilicon-nitride conduction band offset is 2.1 eV, and the effective mass of electrons in the nitride has been assumed to be the same as in the oxide, $0.5 m_e$. These values of the nitride dielectric constant, polysilicon-nitride band offset and effective mass are used in Figs. 6 and 11 as well.

both the single and dual layer gate dielectrics are based on the exact solutions for the tunneling transmission shown, respectively, in Figs. 4(a) and 5(a). The dopant concentrations used in this simulation are 5.0×10^{19} cm⁻³ for polysilicon

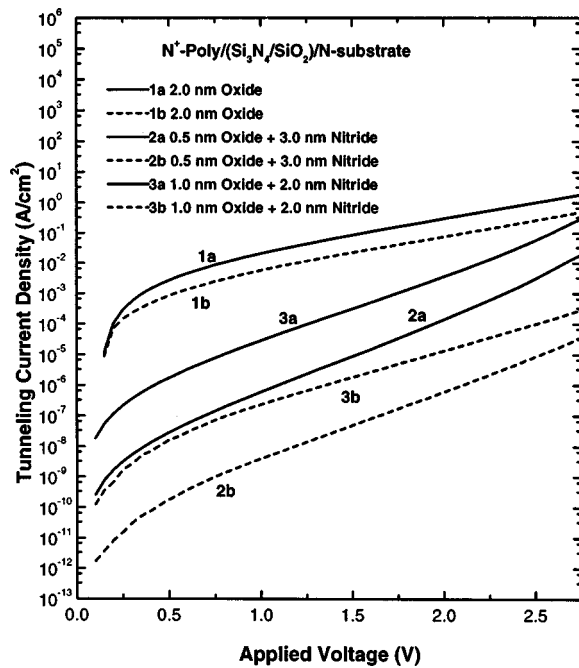


FIG. 6. Comparison between calculated tunneling currents for substrate and gate injection tunneling currents through oxide/nitride double layer gate dielectrics with an equivalent oxide thickness of 2.0 nm with a single oxide layer of thickness 2.0 nm. Solid lines are for substrate injection, and dashed lines for gate injection.

and $1.0 \times 10^{17} \text{ cm}^{-3}$ for n -substrate; these are the same concentrations used for the calculations in Fig. 4(b) and 5(b), and will be used throughout the remainder of this article. The calculations in these two figures demonstrate that physically thicker oxide/nitride dual layer films, having the same capacitance-voltage (or equivalently oxide-equivalent thickness) as the physically thinner single layer oxides, may reduce the tunneling current up to several orders of magnitude for both substrate injection and gate injection for a range of applied bias voltages up to at least 1.5 V. These currents were calculated using the same effective mass for the oxide and nitride regions with the other parameters the same as used in Fig. 4(a). The effective mass for the nitride layer may be smaller than that of the oxide due to the increased dielectric constant and decreased band gap, and the implications of this difference in effective mass between the two layers of the dual gate dielectric structure will be addressed immediately below in this section of the article. However, it is still anticipated that the dual layer structures will have lower direct tunneling currents with respect to single layer oxides with the same oxide equivalent thicknesses. This means from the perspective of direct tunneling currents, these types of oxide/nitride dual layer dielectrics should be useful for future generations of ULSI devices.

To demonstrate the effect that different values of the nitride effective mass have the tunneling current reduction for the dual layer dielectrics with respect to single layer oxides, the tunneling currents shown in Fig. 6 for the 1.0 nm oxide/2.0 nm nitride structures have been plotted in Fig. 7 for nitride effective masses of 1.0 and 0.25. These values of m_N are anticipated to span a range of possible values for nitride dielectrics materials. As expected, reducing the value

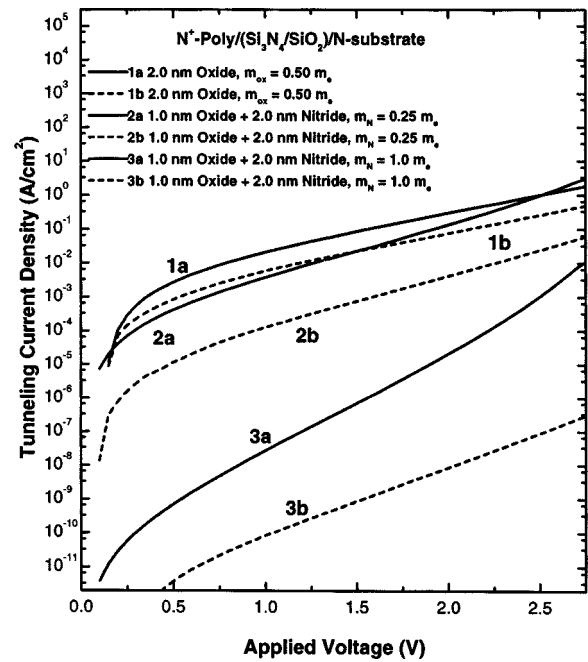


FIG. 7. Simulated results for both substrate and gate injection tunneling currents through 1.0 nm/2.0 nm oxide/nitride double layer gate dielectrics with an equivalent oxide thickness of 2.0 nm and a single oxide layer of thickness 2.0 nm. Here the effective mass for the oxide is $0.5 m_e$. The effective mass for the nitride layer, m_N , is 0.25 for the calculations in Figs. 2(a) and 2(b), and $1.0 m_e$ for those in Figs. 3(a) and 3(b). The solid lines are for substrate injection and the dashed lines for gate injection.

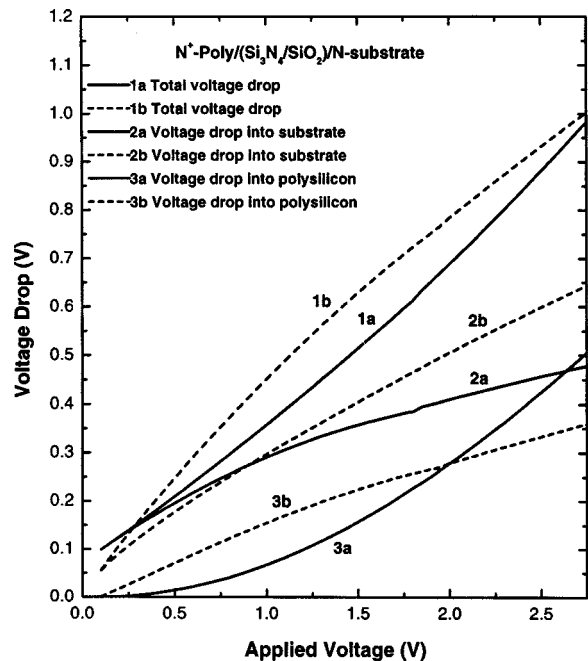


FIG. 8. Potential drops as a function of applied bias voltage in the N -substrate and polysilicon gate electrode for substrate and gate injection for a single oxide layer of thickness 2.0 nm. The total potential drop across the two semiconductors is also displayed. The solid lines are for substrate injection and the dashed lines for gate injection. The doping densities in the n -type substrate and polysilicon gate electrodes are $1.0 \times 10^{17} \text{ cm}^{-3}$ and $5.0 \times 10^{19} \text{ cm}^{-3}$, respectively. These same doping densities are used in Figs. 9, 10, and 11 as well.

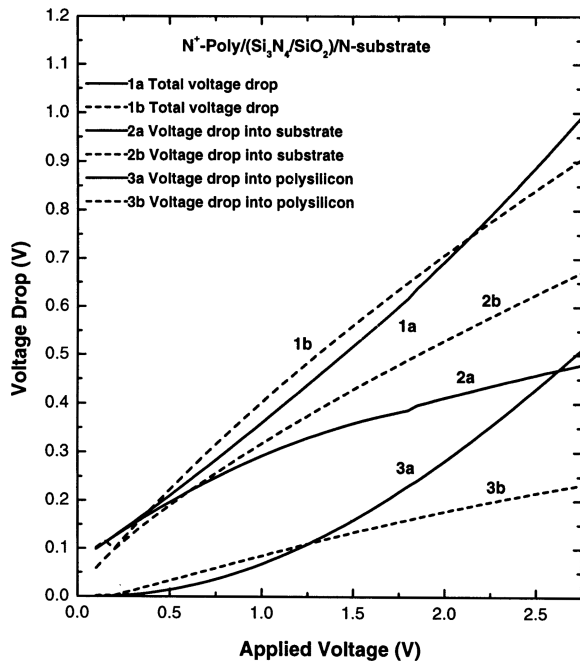


FIG. 9. Potential drops as a function of applied bias voltage in the N-substrate and polysilicon gate electrode for substrate and gate injection for a (0.5 nm/3.0 nm) dual oxide/nitride dielectric with an equivalent oxide thickness 2.0 nm. The total potential drop across the two semiconductors is also displayed. The solid lines are for substrate injection and the dashed lines for gate injection.

of m_N increases the tunneling currents at all voltages, whereas increasing m_N has the opposite effect. The calculations for different values of m_N in Fig. 7 maintain the trends shown in Fig. 6 and demonstrate that tunneling currents in

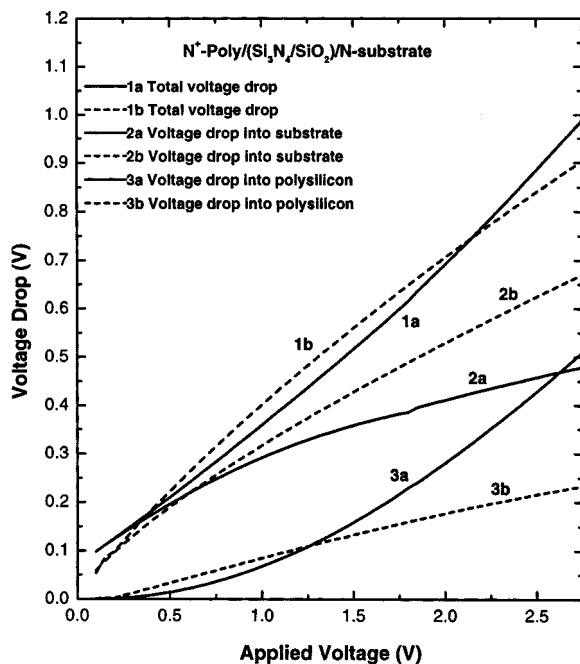


FIG. 10. Potential drops as a function of applied bias voltage in the N-substrate and polysilicon gate electrode for substrate and gate injection for a (1.0 nm/2.0 nm) dual oxide/nitride dielectric with an equivalent oxide thickness 2.0 nm. The total potential drop across the two semiconductors is also displayed. The solid lines are for substrate injection and the dashed lines for gate injection.

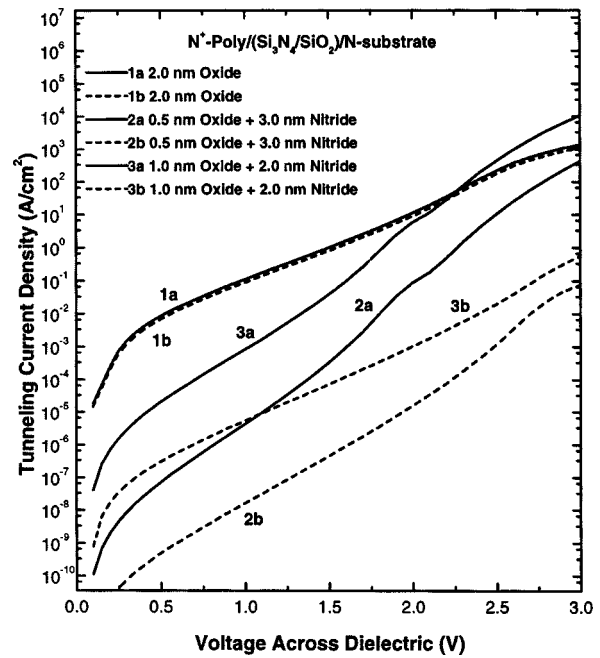


FIG. 11. Calculated tunneling currents as a function of the potential drop across the dielectric layer for substrate and gate injection through oxide/nitride double layer gate dielectrics with an equivalent oxide thickness of 2.0 nm and a single oxide layer of thickness 2.0 nm. The solid lines are for substrate injection and the dashed lines for gate injection.

dual layer dielectrics is generally smaller than for single layer oxide dielectrics, provided the thicknesses of the oxide and nitride layers of the dual layer structures have been chosen on the basis of the relative static dielectric constants to yield the same dielectric or oxide equivalent capacitance.

Figure 8 shows the potential drops in the polysilicon and substrate for both substrate and gate injection for a single oxide layer of thickness 2.0 nm. Similarly, Figs. 9 and 10 present the voltage drops in the polysilicon and substrate for 0.5/3.0 nm oxide/nitride and 1.0/2.0 nm oxide/nitride dual layer films with oxide equivalent thicknesses of 2.0 nm.

Figures 8–10 demonstrate the following: (i) the voltage drop in an inverted substrate for gate injection is larger than for an accumulated substrate for substrate injection; (ii) the voltage drop for a depleted or inverted polysilicon contact for substrate injection is larger than for accumulated polysilicon as in gate injection; (iii) the total voltage drop for a single oxide layer and oxide/nitride dual layer gate dielectrics are not significantly different; and (iv) the total voltage drops for the oxide/nitride 0.5/3.0 nm and 1.0/2.0 nm dual layer films are essentially the same. In our simulations, it has been found that the voltage drop across the polysilicon gate is strongly dependent on the polysilicon doping concentration. For polysilicon concentrations lower than those shown in Figs. 8–10, the potential drop in polysilicon can increase significantly. Similarly, it was also observed that for a given polysilicon concentration, the voltage drop across the polysilicon layer increases as the substrate doping concentrations increases.

It is interesting to plot the tunneling currents for substrate injection and gate injection as a function of the voltage across the dielectric films rather than as a function of the gate

voltages. These results are shown in Fig. 11. Plotted in this way, Fig. 11 demonstrates that the tunneling currents for both substrate injection and gate injection for single layer oxides are essentially the same. Stated differently, this means that electron tunneling for substrate injection and gate injection for single layer oxides is symmetrical. However, electron tunneling for substrate injection and gate injection for oxide/nitride dual layer films shows a large asymmetry. This result can be easily understood: in the WKB approximation, the transmission probability strongly depends on the area under the potential barrier for a given applied bias. From Fig. 2, for substrate injection, and Fig. 3, for gate injection, the asymmetry is clearly evident.

V. CONCLUSIONS

A physical model for obtaining tunneling currents for single oxide layer and oxide/nitride dual layer films for both substrate injection and gate injection has been presented. The following conclusions can be drawn. First, electron tunneling for both single oxide layers and oxide/nitride dual layer films is well described for ultrathin gate dielectrics using the exact solution to tunneling transmission. This is verified by the comparison between the tunneling calculation of the results of Ref. 1, as presented in Fig. 1. As noted earlier in the article, the WKB approximation would give essentially the same result as the exact calculation of the tunneling transmission; however the WKB approximation is inappropriate for the dual layer oxide-nitride structures. For both n -substrate/(oxide)/ n -polysilicon and n -substrate/(oxide/nitride)/ n -polysilicon metal oxide semiconductor structures, the tunneling current for substrate injection is larger than that for gate injection. Preliminary calculations demonstrate increased symmetry for tunneling through ultrathin ONO structures in which the two oxide layers are of the same thickness. This increased symmetry results from increases in gate injection, rather than changes in both gate and substrate injection and therefore is not beneficial in n -MOS transistor structures where off state tunneling is from the gate at zero bias to the positively biased drain contact.

Second, quantization for both accumulation and inversion in the substrate layers is significant and must be taken into account in modeling device operation. In particular, these quantum mechanical effects have been demonstrated to have a significant influence on both direct tunneling and FN tunneling. Third, the effects of polysilicon depletion play an important role in modeling tunneling current. The calculations demonstrate that as the oxide thickness decreases, polysilicon depletion effects become more pronounced as the ratio of the polysilicon doping concentration to the substrate doping concentration decreases. This trend has stimulated interest in re-examining the use of metal gate electrodes. Finally, and perhaps most importantly, the calculations of this article have demonstrated that physically thicker oxide/nitride dual layer films, which have the same capacitance and oxide equivalent thickness as significantly thinner single layer oxides, can produce significant reductions in direct tunneling current. Based on the calculations of this article, dual layer oxide/nitride dielectrics should find applications in

complementary metaloxide semiconductor device structures which require oxide equivalent thicknesses in the range from about 2 to 1.2 nm. The physical thicknesses of these films would typically be about one and a half times greater than the oxide equivalent thicknesses; i.e., from 3 to 1.8 nm. The predictions of these calculations for ON structures require experimental verification. Studies are in progress in our laboratory, wherein ultrathin dual layer structures have been prepared by a combination of low temperature (300 °C) remote plasma processing, and low thermal budget rapid thermal annealing (e.g., 30 s at 900 °C in an inert atmosphere such as Ar).³⁷ Field-effect transistors (FETs) with oxide equivalent thicknesses of ~ 1.8 nm display drive currents that are essentially equal to those of FETs with oxide dielectrics ~ 1.7 nm, but display tunneling currents that are reduced by factors of approximately 5.³⁸ This suggests that the electron effective mass in the nitride layers is less than in the oxide, e.g., $m_n \sim 0.3 m_o$, whereas $m_{ox} \sim 0.5 m_o$ (see Fig. 7).

ACKNOWLEDGMENTS

The authors would like to thank Dr. D. A. Buchanan for providing us his experimental data. The authors acknowledge research support from the Office of Naval Research, the National Science Foundation, and the Semiconductor Research Corporation.

- ¹D. A. Buchanan and S.-H. Lo, *The Physics and Chemistry of SiO₂ and Si-SiO₂ Interface*, edited by H. Z. Massoud, E. H. Poindexter, and C. R. Helms (The Electrochemical Society, Pennington, NJ, 1996), pp. 319–322.
- ²C. Hu, IEDM, 319 (1996).
- ³H. Wong and Y. C. Cheng, J. Appl. Phys. **70**, 1078 (1991).
- ⁴X. R. Cheng, Y. C. Cheng, and B. Y. Liu, J. Appl. Phys. **63**, 797 (1988).
- ⁵H. Fang, H. S. Krisch, B. J. Gross, C. G. Sodini, J. Chung, and D. A. Antoniadou, IEEE Electron Device Lett. **EDL-13**, 217 (1992).
- ⁶Y. Ma, T. Yasuda, and G. Lucovsky, Appl. Phys. Lett. **64**, 2226 (1994).
- ⁷T. Ito, T. Nakamura, and H. Ishikawa, IEEE Electron Device Lett. **EDL-29**, 498 (1982).
- ⁸K. K. Young, C. Hu, and W. G. Oldham, Appl. Surf. Sci. **30**, 171 (1987).
- ⁹S. Mori, E. Sakagami, H. Araki, Y. Kaneeko, K. Narita, Y. Ohshima, N. Arai, and K. Yoshikama, IEEE Electron Device Lett. **EDL-38**, 386 (1991).
- ¹⁰G. W. Yoon, J. Kim, L. K. Han, J. Yan, and D. L. Kwong, Proc. Symp. VLSI Tech. Dig. (1994), p. 155.
- ¹¹L. D. Yau, IEEE Electron Device Lett. **EDL-7**, 365 (1986).
- ¹²R. Baunach and A. Spitzer, Appl. Surf. Sci. **30**, 180 (1987).
- ¹³S. Manzini and G. Queirolo, Solid-State Electron. **30**, 587 (1987).
- ¹⁴K. K. Young, C. Hu, and W. G. Oldham, IEEE Electron Device Lett. **EDL-9**, 616 (1988).
- ¹⁵A. Spitzer and R. Baunach, Appl. Surf. Sci. **39**, 200 (1989).
- ¹⁶L. Dori, J. Sun, M. Arienzo, S. Basavaiah, Y. Taur, and D. Zichermann, Proc. Symp. VLSI Tech. Dig. (1987) p. 25.
- ¹⁷H. S. Momose, T. Morimoto, Y. Ozawa, K. Yamchi, and H. Iwai, IEEE Trans. Electron Devices **ED-42**, 701 (1995).
- ¹⁸W. A. Harrison, Phys. Rev. **123**, 85 (1961).
- ¹⁹N. G. Tarr, D. L. Pulfrey, and D. S. Camporese, IEEE Trans. Electron Devices **ED-30**, 1780 (1983).
- ²⁰M. Depas, B. Uermeire, P. W. Mertens, R. L. Van Meirnaeghe, and M. M. Hegus, Solid-State Electron. **38**, 1465 (1994).
- ²¹E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), Chap. 7.
- ²²C. B. Duke, Phys. Rev. **159**, 632 (1967).
- ²³J. A. Appabbaum and G. A. Barat, Phys. Rev. B **4**, 1225 (1971).
- ²⁴F. Stern, Phys. Rev. **135**, 4891 (1972).
- ²⁵Z. A. Weinburg, Solid-State Electron. **20**, 11 (1977).
- ²⁶T. Ando, A. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
- ²⁷J. Sune, P. Olivo, and B. Ricco, J. Appl. Phys. **1**, 337 (1991).
- ²⁸J. Sune, P. Olivo, and B. Ricco, IEEE Trans. Electron Devices **ED-39**, 1732 (1992).

- ²⁹F. Rana, S. Tiwari, and D. A. Buchanan, *Appl. Phys. Lett.* **69**, 1104 (1996).
- ³⁰E. Rosenbaum, R. Moazzami, and C. Hu, Proceedings of the International Symposium on Very Large Scale Integrated Tech., Syst. and Applications, 1991 (unpublished), p. 214.
- ³¹K. F. Schuegraf, C. C. King, and C. Hu, Proceedings of the International Symposium on Very Large Scale Integrated Tech., Syst. and Applications, 1993 (unpublished), p. 83.
- ³²K. F. Schuegraf and C. Hu, *IEEE Trans. Electron Devices* **ED-41**, 761 (1994).
- ³³S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981), p. 19.
- ³⁴M. Shur, *Physics of Semiconductor Devices* (Prentice-Hall, Englewood Cliffs, NJ, 1990), p. 47.
- ³⁵H. J. Wen, R. Ludeke, D. M. Newns, and S. H. Lo, *J. Vac. Sci. Technol. A* **15**, 784 (1997).
- ³⁶D. A. Buchanan, SRC Topical Research Conference on Ultrathin Gate Dielectrics—Technology, Reliability and Characterization, edited by H. Z. Massoud, D. A. Buchanan, and V. Bissessur, 1996 (unpublished).
- ³⁷G. Lucovsky *et al.*, *The Physics and Chemistry of SiO₂ and Si-SiO₂ Interface*, edited by H. Z. Massaud, E. H. Poindexter, and C. R. Helms (The Electrochemical Society, Pennington, NJ, 1996), p. 441.
- ³⁸C. R. Parker, G. Lucovsky, and J. R. Hauser, *IEEE Electron Dev. Lett.* (submitted).