Analytic model for direct tunneling current in polycrystalline silicon-gate metal–oxide–semiconductor devices

Leonard F. Register, Elyse Rosenbaum, and Kevin Yang

Citation: Appl. Phys. Lett. **74**, 457 (1999); doi: 10.1063/1.123060
View online: https://doi.org/10.1063/1.123060
View Table of Contents: http://aip.scitation.org/toc/apl/74/3
Published by the American Institute of Physics

Analytic model for direct tunneling current in polycrystalline silicon-gate metal–oxide–semiconductor devices

Leonard F. Register[a)]
Beckman Institute and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Elyse Rosenbaum and Kevin Yang[b)]
Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 30 June 1998; accepted for publication 10 November 1998)

An analytic model of the direct tunneling current in metal–oxide–semiconductor devices as a function of oxide field is presented. Accurate modeling of the low-field roll-off in the current results from proper modeling of the field dependencies of the sheet charge, electron impact frequency on the interface, and tunneling probability. To obtain the latter dependence, a modified WKB approximation is used. © *1999 American Institute of Physics.* [S0003-6951(99)02303-7]

Conventional transistor scaling theory requires that the gate oxide thickness t_{ox} be reduced as the minimum channel length is scaled down. However, for $t_{ox}<4$ nm, the gate current in metal–oxide–semiconductor (MOS) devices is significant even for oxide potential drops $|V_{ox}|$ substantially less than the Si–SiO_2 conduction band discontinuity $\chi_c \simeq 3.15$ V due to so-called ''direct tunneling'' of electrons through the oxide. Gate oxide thickness in commercial ICs is rapidly approaching the limit where direct tunneling would lead to an unacceptable increase in the off-state power consumption.[1,2] It has also been projected that the power supply voltage for high performance microprocessors will be reduced to 1.2 V, or less, within the next decade.[1] Therefore, accurate modeling of direct tunneling currents at low voltages is essential.

Initial attempts to model the direct tunneling current analytically provided good fit to data for $1\ \text{V} \leqslant |V_{ox}| \leqslant \chi_c$, but error approached three orders of magnitude as V_{ox} approached 0 V.[3,4] Improved fit at low voltages/fields was obtained by employing more rigorous and computationally intensive models.[5] Recently, excellent agreement was obtained throughout the direct tunneling regime for several oxide thicknesses via self-consistent numerical solution of Schrödinger's and Poisson's equations, and rigorous modeling of tunneling from bound states.[2]

In this letter, a wholly analytic model of direct tunneling is presented that combines the convenience of the analytic methods of Refs. 3 and 4 with accuracy comparable to the computationally intensive method of Ref. 2. This result is achieved by the careful use of simplifying approximations that do not compromise the essential physics of the analysis. Accurate modeling of the low-field rolloff in the current results from proper modeling of the field dependencies of the sheet charge Q, electron impact frequency on the interface f, and mean tunneling probability $\overline{T}$. The gate current J_g is then given in terms of these quantities simply as $J_g = QfT$.

The basic features of this model are illustrated in Fig. 1. In accumulation, quantum confinement normal to the interface results in splitting of the conduction band into subbands.[2,5,6] To facilitate analysis, only the lowest lying subband is considered. The dependence of the subband quantization energy $E_{si,\perp}$ on the magnitude of the field within the oxide F_{ox} is approximated by modeling the potential well as triangular with slope $dE_c/dx = -q(\epsilon_{ox}/\epsilon_{si})F_{ox}$ where ϵ_{si} and ϵ_{ox} are the silicon and oxide dielectric constants, respectively, and then requiring $2\int_{x_c}^{0} k_{si,\perp}(x) = 2\pi$ where x_c is the classical turning point in the silicon and $k_{si,\perp}$ is the interface–normal component of the electron wave vector. As in Ref. 5, the impact frequency f of each electron on the Si–SiO_2 interface is approximated by $f^{-1} = 2\int_{x_c}^{0} v_{si,\perp}^{-1}(x)\,dx$, where $v_{si,\perp}$ is the interface–normal component of electron group velocity. Modeling the energy band as parabolic, i.e.,

$$k_{si,\perp} = \sqrt{\frac{2m_{si,\perp}E_{si,\perp}}{\hbar^2}}, \tag{1}$$

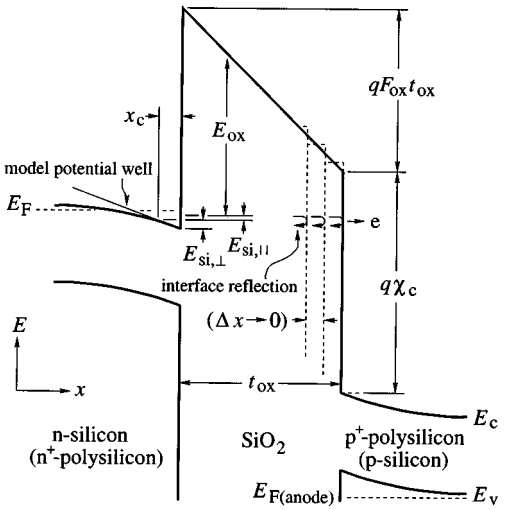

FIG. 1. Band diagram for direct tunneling from accumulated n-Si (n^+-poly-Si) to accumulated p^+-poly-Si (p-Si). The accumulation well is approximated as triangular. The electron wave function is partially reflected at each potential discontinuity in addition to being attenuated between discontinuities.

a)Electronic mail: register@ceg.uiuc.edu
b)Present address: Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

$$v_{\mathrm{si},\perp}=\sqrt{\frac{2E_{\mathrm{si},\perp}}{m_{\mathrm{si},\perp}}}, \tag{2}$$

where $m_{\mathrm{si},\perp}$ is the interface–normal component of effective mass, this analysis yields

$$E_{\mathrm{si},\perp}=0.6\times\frac{(3\pi\hbar qm_{\mathrm{si},\perp})^{2/3}}{2m_{\mathrm{si},\perp}}\left(\frac{\varepsilon_{\mathrm{ox}}F_{\mathrm{ox}}}{\varepsilon_{\mathrm{si}}}\right)^{2/3}, \tag{3}$$

and

$$f=0.6\times\frac{2q}{(3\pi\hbar qm_{\mathrm{si},\perp})^{1/3}}\left(\frac{\varepsilon_{\mathrm{ox}}F_{\mathrm{ox}}}{\varepsilon_{\mathrm{si}}}\right)^{2/3}. \tag{4}$$

The leading factor of 0.6 in Eqs. (3) and (4) compensates approximately for the exaggerated electron confinement of this model independent of doping concentration, as measured by comparison to numerical analysis.[6]

The charge available for tunneling Q is modeled as field induced, i.e.,

$$Q=\epsilon_{\mathrm{ox}}F_{\mathrm{ox}}. \tag{5}$$

Relating this sheet charge density to the number of occupied subband states gives the Fermi energy E_{F}:

$$E_{\mathrm{F}}-E_{\mathrm{si},\perp}=\frac{\pi\hbar^2}{qm_{\mathrm{si},\parallel}\eta}\epsilon_{\mathrm{ox}}F_{\mathrm{ox}}, \tag{6}$$

where $m_{\mathrm{si},\parallel}/(\pi\hbar^2)$ is the density of states, ''$\parallel$'' indicates the interface parallel component, η is the number of orientationally equivalent energy valleys, and the Fermi distribution has been approximated by a step function.

To calculate the tunneling probabilities, a modified WKB approach is used:

$$T=T_{\mathrm{WKB}}T_{\mathrm{R}}, \tag{7}$$

where T_{WKB} is the usual WKB tunneling probability valid for smoothly varying potentials, and T_{R} corrects for reflections from potential discontinuities.[7] As in Ref. 2, based on experimental evidence,[8,9] the barrier height to tunneling is taken to be a function only of the total electron energy, and the SiO_2 band-gap dispersion relation is modeled as a two-band Franz-type,[2,10] i.e.,

$$\frac{\hbar^2\kappa_{\mathrm{ox}}^2}{2m_{\mathrm{ox}}}=\gamma=E_{\mathrm{ox}}\left(1-\frac{E_{\mathrm{ox}}}{E_{\mathrm{g}}}\right) \tag{8}$$

$$\nu_{\mathrm{ox}}=\frac{1}{\gamma'}\sqrt{\frac{2\gamma}{m_{\mathrm{ox}}}}, \tag{9}$$

where

$$\gamma'=\frac{d\gamma}{dE_{\mathrm{ox}}}=\left(1-\frac{2E_{\mathrm{ox}}}{E_{\mathrm{g}}}\right). \tag{10}$$

Here, κ_{ox} and ν_{ox} are the magnitudes of the imaginary wave vector and group velocity, respectively. E_{ox} is the magnitude of the electron energy referenced to the oxide conduction band edge, m_{ox} is the band-edge effective mass, and E_{g} is the oxide band gap $\simeq 9$ eV.[11] Under these conditions,

$$T_{\mathrm{WKB}}=\exp\left[\frac{E_{\mathrm{g}}\sqrt{2m_{\mathrm{ox}}}}{4\hbar qF_{\mathrm{ox}}}(2\gamma'\sqrt{\gamma}+\sqrt{E_{\mathrm{g}}}\sin^{-1}\gamma')\Big|_{E_{\mathrm{ox}}=q\phi_{\mathrm{an}}}^{E_{\mathrm{ox}}=q\phi_{\mathrm{cat}}}\right], \tag{11}$$

where

$$q\phi_{\mathrm{cat}}=q\chi_{\mathrm{c}}-(E_{\mathrm{si},\perp}+E_{\mathrm{si},\parallel}), \tag{12}$$

$$q\phi_{\mathrm{an}}=q\chi_{\mathrm{c}}-(E_{\mathrm{si},\perp}+E_{\mathrm{si},\parallel})-qF_{\mathrm{ox}}t_{\mathrm{ox}}, \tag{13}$$

are the net barrier heights for the electron at the cathode and anode interfaces, respectively. The correction factor has the band-structure independent form,[7]

$$T_{\mathrm{R}}=\frac{4v_{\mathrm{si},\perp}(E_{\mathrm{si},\perp})\nu_{\mathrm{ox}}(q\phi_{\mathrm{cat}})}{v_{\mathrm{si},\perp}^2(E_{\mathrm{si},\perp})+\nu_{\mathrm{ox}}^2(q\phi_{\mathrm{cat}})}\times\frac{4v_{\mathrm{si},\perp}(E_{\mathrm{si},\perp}+qF_{\mathrm{ox}}t_{\mathrm{ox}})\nu_{\mathrm{ox}}(q\phi_{\mathrm{an}})}{v_{\mathrm{si},\perp}^2(E_{\mathrm{si},\perp}+qF_{\mathrm{ox}}t_{\mathrm{ox}})+\nu_{\mathrm{ox}}^2(q\phi_{\mathrm{an}})}. \tag{14}$$

T_{R} is obtained by considering reflections from the material interfaces and, essential to the analysis, the model interfaces internal to the oxide are shown in Fig. 1 with $\Delta x\rightarrow 0$. Multiple reflections within any Δx region are neglected, an approximation that eventually fails close to the transition from direct to Fowler–Nordheim tunneling.

With the relation of accumulated charge, impact frequency and tunneling probability to oxide field is established, and the gate current can be evaluated analytically:

$$J_{\mathrm{g}}=\frac{\eta qm_{\mathrm{si},\parallel}f}{\pi\hbar^2}\int_0^{E_{\mathrm{F}}-E_{\mathrm{si},\perp}}TdE_{\mathrm{si},\parallel}\simeq QfT|_{\mathrm{E_{si,\parallel}=1/2(E_f-E_{si,\perp})}}. \tag{15}$$

In Eq. (15), a midpoint approximation to the integral is used for simplicity to obtain the mean tunneling probability while introducing very little error.

Model predictions of J_{g} vs F_{ox} were compared with data for both p^+ polysilicon/oxide/n-silicon and n^+-polysilicon/oxide/p-silicon devices. The p^+-polysilicon gate devices have conventional SiO_2 gate oxide; the gate oxide for the n^+-polysilicon gate devices was grown at 750 °C and annealed in NO.[12] The $V_{\mathrm{g}}-F_{\mathrm{ox}}$ relationships were obtained by integration of C–V data.[13] The oxide thicknesses were measured by TEM calibrated ellipsometry to within an uncertainty of ± 0.15 nm. To provide the best overall fit to data, m_{ox} was set equal to $0.61\,m_{\mathrm{e}}$. Notably, in the Fowler–Nordheim tunneling regime, the Franz-type dispersion relation with this band-edge effective mass is equivalent within the WKB approximation to a parabolic dispersion relation with an effective mass of $0.50\,m_{\mathrm{e}}$, the value found in the well known work of Weinberg.[8] The small difference between the band-edge mass used here and that used in the numerical work of Ref. 2, $0.55\,m_{\mathrm{e}}$, is attributed to the approximate quantum treatments used here to retain analytic tractability, and to the uncertainty in the measured oxide thicknesses.

Figure 2 shows measured and predicted J_{g} for a p^+-polysilicon/oxide/n-silicon device with $t_{\mathrm{ox}}\simeq 2.5$ nm under positive gate bias. In the $\langle 100\rangle$ oriented cathode, $\eta=2$, $m_{\mathrm{si},\perp}=0.98\,m_{\mathrm{e}}$, and $m_{\mathrm{si},\parallel}=0.19\,m_{\mathrm{e}}$, where m_{e} is the free-

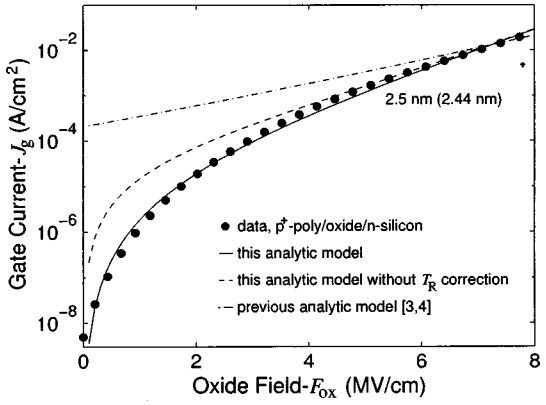

FIG. 2. Measured and model gate current densities J_g vs oxide field F_{ox}. The nominal oxide thickness from calibrated ellipsometry was 2.5 nm; the oxide thickness used for the model, 2.44 nm, is well within the range of measurement error.

space electron mass. This figure also shows the effects of neglecting the tunneling correction factor T_R in this model, and results for the previous analytic model in which the cathode was modeled as metal-like, i.e., as being a field-independent source of electrons for tunneling.[3,4] In the latter case, to provide the best possible fit, a parabolic dispersion relation within the oxide was assumed with $m_{ox}=0.38\,m_e$.

Measured and predicted J_g for three different oxide thicknesses are compared in Fig. 3. The two thinnest samples were n^+-polysilicon/oxide/p-silicon capacitors characterized

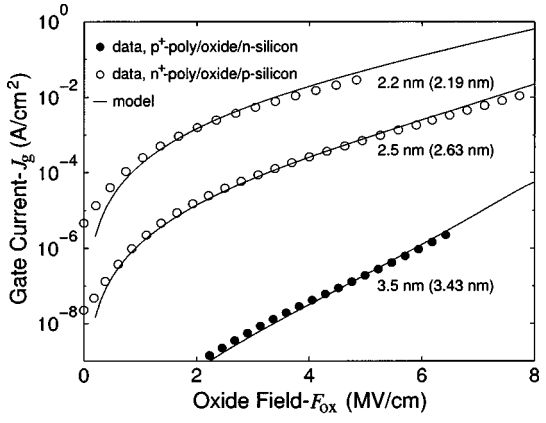

FIG. 3. J_g vs F_{ox} for three oxide thicknesses. Nominal t_{ox} measured by ellipsometry is noted in the figure; values in parentheses provide the best fit of the model to data. All devices were biased in accumulation.

under negative gate bias. Despite heavy doping in the polysilicon cathodes, the only adjustments in the model were those of setting $m_{si,\perp}=0.32m_e$, $m_{si,\parallel}=0.25m_e$, and $\eta=4$ to approximate the band-structure for the predominantly $\langle 110\rangle$ oriented silicon grains.

In conclusion, an analytic model for the direct tunneling current in MOS devices as a function of field that shows very good fit to data has been developed. Consideration of the field dependencies of the charge Q, impact frequency f, and WKB correction factor T_R results in accurate modeling of the low-field roll-off in the current despite simplifying approximations used to retain analytic tractability. Although only tunneling from accumulation layers was examined in this letter, it also should be possible to model tunneling from inversion layers by subtracting the field required to achieve strong inversion from F_{ox} in Eqs. (5) and (6).

The authors thank Erhong Li for device measurements and Hewlett Packard Laboratories and Advanced Micro Devices for donating the test devices. In particular, we wish to acknowledge Mike Cox and Carlos Dìaz of the ULSI Research Lab, Hewlett-Packard Laboratories, Palo Alto, CA, and Peng Fang and Jon Tao of Advanced Micro Devices, Sunnyvale, CA for device fabrication. This work is supported by the Joint Services Electronics Program under Contract No. N00014-96-1-0219, the National Science Foundation, and the Semiconductor Research Corporation.

[1] Semiconductor Industry Association, National Technology Roadmap for Semiconductors (1997).
[2] S.-H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, IEEE Electron Device Lett. **18**, 209 (1997).
[3] K. F. Schuegraf, C. C. King, and C. Hu, Symp. VLSI Technol. Digest of Tech. Papers, 18, (1992); K. F. Schuegraf and C. Hu, IEEE Trans. Electron Devices **41**, 761 (1994).
[4] M. Depas, B. Vermeire, P. W. Mertens, R. L. Van Meirhaeghe, and M. M. Heyns, Solid-State Electron. **38**, 1465 (1995).
[5] F. Rana, S. Tiwari, and D. A. Buchanan, Appl. Phys. Lett. **69**, 1104 (1996).
[6] J. Sune, P. Olivo, and B. Ricco, IEEE Trans. Electron Devices **39**, 1732 (1992).
[7] L. F. Register (unpublished).
[8] Z. A. Weinberg, J. Appl. Phys. **53**, 5052 (1982).
[9] Z. A. Weinberg and A. Harstein, J. Appl. Phys. **54**, 2517 (1983).
[10] J. Maserjian and G. P. Petersson, Appl. Phys. Lett. **25**, 50 (1974).
[11] S. M. Sze, *Physics of Semiconductor Devices* (Wiley-Interscience, New York, 1981).
[12] C. Dìaz, M. Cox, W. Greene, F. Perlaki, E. Carr, I. Manna, A. Bayoumi, M. Cao, N. Shamma, M. Tavassoli, C. Chi, N. Farrar, D. Lefforge, Y. Chang, B. Langley, and P. Marcoux, VLSI Technol. Symp. (1997).
[13] E. Rosenbaum and L. F. Register, IEEE Trans. Electron Devices **44**, 317 (1997).