Citation for published version:
Slavcheva, G, Davies, JH, Brown, AR & Asenov, A 2002, 'Potential fluctuations in metal-oxide-semiconductor field-effect transistors generated by randomly distributed impurities in the depletion layer', *Journal of Applied Physics*, vol. 91, no. 7, pp. 4326-4334 . https://doi.org/10.1063/1.1450031

DOI:
10.1063/1.1450031

Publication date:
2002

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Copyright 2002 American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics.

The following article appeared in Slavcheva, G, Davies, JH, Brown, AR & Asenov, A 2002, 'Potential fluctuations in metal-oxide-semiconductor field-effect transistors generated by randomly distributed impurities in the depletion layer' Journal of Applied Physics, vol 91, no. 7, pp. 4326-4334 and may be found at http://dx.doi.org/10.1063/1.1450031

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 19. May. 2024

JOURNAL OF APPLIED PHYSICS VOLUME 91, NUMBER 7 1 APRIL 2002

Potential fluctuations in metal–oxide–semiconductor field-effect transistors generated by random impurities in the depletion layer

G. Slavcheva, J. H. Davies,[a] A. R. Brown, and A. Asenov
Device Modelling Group, Department of Electronics and Electrical Engineering, University of Glasgow, Glasgow G12 8QQ, United Kingdom

(Received 1 October 2001; accepted for publication 17 December 2001)

We have studied the magnitude and length scale of potential fluctuations in the channel of metal–oxide–semiconductor field-effect transistors due to the random positions of ionized impurities in the depletion layer. These fluctuations effect the threshold voltage of deep submicron devices, impede their integration, and reduce yield and reliability. Our simple, analytic results complement numerical, atomistic simulations. The calculations are based on a model introduced by Brews to study fluctuations due to charges in the oxide. We find a typical standard deviation of 70 mV in the potential below threshold, where the channel is empty, falling to 40 mV above threshold due to screening by carriers in the channel. These figures can be reduced by a lightly doped epitaxial layer of a few nm thickness. The correlation function decays exponentially in an empty channel with a length scale of 9 nm, which screening by carriers reduces to about 5 nm. These calculations of the random potential provide a guide to fluctuations of the threshold voltage between devices because the length of the critical region in a well-scaled transistor near threshold is comparable to the correlation length of the fluctuations. The results agree reasonably well with atomistic simulations but detailed comparison is difficult because half of the total standard deviation comes from impurities within 1 nm of the silicon–oxide interface, which is a single layer of the grid used in the simulations. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1450031]

I. INTRODUCTION

The discreteness and randomness of ionized impurities near the channel introduce strong fluctuations into the characteristics of metal–oxide–semiconductor field-effect transistors (MOSFETs) when they are scaled to deep submicron dimensions. They also reduce the average threshold voltage because current can percolate through the favorable regions of the fluctuations in the potential in the channel. These effects, predicted twenty years ago,[1,2] have been confirmed in several experiments[3–5] and simulations.[6–10] The fluctuations will have significant impact on the functionality, yield, and reliability of the corresponding systems.

Most random charges in early devices arose from the oxide and the silicon–oxide interface. The impact of these charges on the interface potential,[11–16] carrier density,[17] and mobility[18] have been studied. However, improved technology has reduced the density of charges in the oxide, while scaling laws require greatly increased doping in the channel of deep submicron devices. Thus, the dominant source of random charges is now the depletion layer near the channel rather than the oxide, a situation that also holds in III–V devices.[19,20]

Our aim in this article is to calculate statistics of fluctuations in the potential in the channel of a MOSFET due to random dopants in the depletion layer. Both the standard deviation and correlation length are needed to estimate fluctuations in the threshold voltage. For example, the standard deviation of the threshold voltage is the same as that of the random potential if the transistor is small compared with the correlation length. In the opposite limit, a breakdown of the real device into a mosaic of statistically different transistors[21] may be an appropriate model if the device is large compared with the correlation length; this is close to the checkerboard considered by Keyes.[2]

We have derived approximate, analytic formulas for the standard deviation of the random potential in the limit of an empty channel, when there is no screening by carriers, and a full channel, where screening by the carriers is linear. A typical standard deviation is around 70 mV for an empty channel, falling to about 40 mV with carriers. The fluctuations may be reduced dramatically by leaving a thin, undoped spacer layer near the interface; a spacer of 4 nm halves the standard deviation. We have also calculated the correlation length of the fluctuations in the channel, which is typically around half the thickness of the depletion region in a well-scaled device below threshold. The standard deviation of the threshold voltage between devices is close to that of the random potential under these conditions. The results agree well with numerical evaluation of the full analytic formulas and with atomistic simulations of deep submicron devices.

Our calculations are based on the simple model illustrated in Fig. 1. We take z as the normal to the interfaces of the MOSFET.

(1) The gate is a perfect conductor and occupies the region $z < -d$.
(2) The oxide has thickness d and dielectric constant κ_{ox}. Both fluctuations in its thickness and random charges are neglected.

[a]Electronic mail: jdavies@elec.gla.ac.uk

 © 2002 American Institute of Physics

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

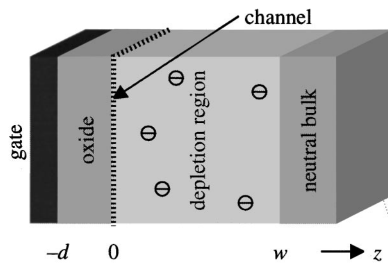

FIG. 1. Sketch of a MOSFET showing gate, oxide, and random charges in depletion layer.

(3) The channel has infinitesimal thickness and lies at the silicon–oxide interface in the plane $z=0$.
(4) The depletion layer, of thickness w and dielectric constant κ_{sc}, contains no carriers. The acceptors are fully ionized and random in x and y with average concentration $N(z)$, which may vary as a function of depth.
(5) The bulk of semiconductor ($z>w$) is neutral; its interface with the depletion layer is abrupt and flat, and is therefore treated like a metal plate at $z=w$.[12]

Screening by electrons in channel will be treated in two limits.

(1) Far below threshold, the channel is taken to be empty and there is no screening at all.
(2) Well above threshold, the density of electrons is so high that linear screening holds.

The practical definition of threshold[9] is closer to the first limit. We do not consider the situation just above threshold, where the density of electrons is grossly inhomogeneous.

We shall first review the Green's function for the electrostatic potential, which is central to the calculations. This is used to construct the power spectrum of the fluctuations, integrated to give the variance, and Fourier transformed for the autocorrelation function. Finally, the results will be compared with atomistic simulations. As an example we take a device with oxide $d=3$ nm thick, depletion layer $w=18$ nm thick, doping of channel $N_A=5\times10^{24}\ \mathrm{m}^{-3}$, and density of electrons above threshold $N_i=10^{16}\ \mathrm{m}^{-2}$. The dielectric constants are $\kappa_{sc}=11.8$ for the semiconductor and $\kappa_{ox}=3.9$ for the oxide.

II. GREEN'S FUNCTION

The central ingredient of the calculations is the Green's function $G(|\mathbf{r}-\mathbf{r}'|;z,z')$, which gives the electrostatic potential at $(\mathbf{r},z)$ due to a unit charge at $(\mathbf{r}',z')$. We have assumed translational symmetry in the plane $\mathbf{r}=(x,y)$. For most calculations, it is more convenient to use the two-dimensional Fourier transform $\tilde{G}(q;z,z')$, defined by

$$\tilde{G}(q;z,z')=\int G(r;z,z')e^{-i\mathbf{q}\cdot\mathbf{r}}d^2\mathbf{r}. \tag{2.1}$$

The Coulomb potential is screened by three effects in a MOSFET. First, the dielectric constants of the silicon and oxide are always present in the background. The second effect is geometrical screening due to the boundaries set by the gate and the edge of the depletion layer. Finally, free carriers in the channel give further screening above threshold. We shall consider the last two effects separately and in combination.

A. Boundaries alone

For both z and z' in the depletion layer, $0\leqslant z\leqslant w$, the Green's function for an empty channel in a region bounded by the gate and edge of the depletion layer is[12]

$$\tilde{G}_0(q;z,z')=\frac{\sinh q(w-z_>)}{\epsilon_0\kappa_{SC}q}\times\frac{\kappa_{ox}\cosh qd\sinh qz_<+\kappa_{sc}\sinh qd\cosh qz_<}{\kappa_{ox}\sinh qw\cosh qd+\kappa_{sc}\sinh qd\cosh qw}, \tag{2.2}$$

where $z_<$ and $z_>$ are the lesser and greater of z and z'. The point of observation will always lie in the silicon–oxide interface, $z=0$, and the random charges have $z\geqslant0$ so we can simplify this result to

$$\tilde{G}_0(q;0,z)=\frac{\sinh q(w-z)}{\sinh qw}\tilde{G}_0(q;0,0), \tag{2.3}$$

where the Green's function for both points in the channel is given by

$$[\tilde{G}_0(q;0,0)]^{-1}=\epsilon q(\kappa_{ox}\coth qd+\kappa_{sc}\coth qw). \tag{2.4}$$

This shows the separate contributions of the finite slabs of oxide and semiconductor. In the limit of very thick slabs $\tilde{G}_0(q;0,0)=1/2\epsilon_0\bar{\kappa}q$, the result for a free charge in a medium of average dielectric constant $\bar{\kappa}=1/2(\kappa_{ox}+\kappa_{sc})$. Brews[12] derived the useful and accurate approximation

$$[\tilde{G}_0(q;0,0)]^{-1}\approx2\epsilon_0\bar{\kappa}\sqrt{Q_s^2+q^2}=\sqrt{(C_{ox}+C_{sc})^2+(2\epsilon_0\bar{\kappa}q)^2}, \tag{2.5}$$

where the geometrical screening wave vector is

$$Q_s=\frac{\kappa_{ox}/d+\kappa_{sc}/w}{\kappa_{ox}+\kappa_{sc}}=\frac{C_{ox}+C_{sc}}{2\epsilon_0\bar{\kappa}}, \tag{2.6}$$

and the capacitance per unit area of the oxide and depletion layers are $C_{ox}=\epsilon_0\kappa_{ox}/d$ and $C_{sc}=\epsilon_0\kappa_{sc}/w$. The wave vector $Q_s\approx0.12\ \mathrm{nm}^{-1}$ here; the oxide usually makes the greater contribution but only by a factor of 2 or so. This approximate Green's function can be inverted to real space, giving

$$G_0(r;0,0)\approx\frac{\exp(-Q_sr)}{4\pi\epsilon_0\bar{\kappa}r}. \tag{2.7}$$

This is a screened Coulomb potential decaying exponentially with the characteristic length $1/Q_s\approx8$ nm. It shows that long-ranged potential fluctuations are damped by the image charges induced in the gate and the edge of the depletion layer. Curves (a) in Fig. 2 shows the exact and approximate Green's functions, which are very close indeed.

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

B. Carriers alone

Now consider a high density N_i of carriers in the channel, which gives linear screening, but assume that the oxide and depletion layers are very thick. In this case, the boundary conditions at $z=-d$ and $z=w$ can be neglected. Linear screening can be represented as a further capacitance $C_i = e^2(dN_i/dE_F)$ in the Green's function, where the derivative is the thermodynamic density of states. At low temperature, this is the conventional density of states at the Fermi level, while $dN_i/dE_F = N_i/2k_BT$ at high temperature.[17] The screening wave vector due to the inversion layer is given by $q_s = (e^2/2\epsilon_0\kappa)(dN_i/dE_F) = C_i/2\epsilon_0\bar{\kappa} \approx 0.22\ \mathrm{nm}^{-1}$ at room temperature for our example. The Green's function is given by[22]

$$\tilde{G}_i(q;0,0) = \frac{1}{2\epsilon_0\bar{\kappa}(q_s+q)}. \tag{2.8}$$

The expression in real space is complicated[23] but its asymptotic decay is

$$G_i(r;0,0) \sim \frac{1}{4\pi\epsilon_0\bar{\kappa}q_s^2 r^3}. \tag{2.9}$$

This inverse-cube decay with distance [curves (b) in Fig. 2] is much slower than the exponential decay found in both three-dimensional screening due to free carriers and the two-dimensional system with boundaries alone.

C. Boundaries and carriers

For general z and z' within the depletion layer the Green's function including screening by both boundaries and carriers is given by

$$\tilde{G}(q;z,z') = \tilde{G}_0(q;z,z') - \frac{\tilde{G}_0(q;z,0)C_i\tilde{G}_0(q;0,z')}{1+C_i\tilde{G}_0(q;0,0)}. \tag{2.10}$$

This is formally similar to the Green's function for the energy level of a short-ranged impurity.[24] If the two distances obey $z \leqslant 0 \leqslant z'$ or $z' \leqslant 0 \leqslant z$, which includes the case of interest, the identity $\tilde{G}_0(q;z,0)\tilde{G}_0(q;0,z') = \tilde{G}_0(q;0,0)\tilde{G}_0(q;z,z')$ holds and the Green's function simplifies to

$$\tilde{G}(q;z,z') = \frac{\tilde{G}_0(q;z,z')}{1+C_i\tilde{G}_0(q;0,0)}. \tag{2.11}$$

The denominator of this can be regarded as an effective dielectric constant due to the carriers in the channel, constrained by the metal plates, and is given by

$$\epsilon_{\mathrm{eff}}(q) = 1 + \frac{q_s}{q}\left[\frac{\kappa_{\mathrm{ox}}\coth qd + \kappa_{\mathrm{sc}}\coth qw}{2\bar{\kappa}}\right]^{-1}. \tag{2.12}$$

This reduces to the usual result $\epsilon_{\mathrm{eff}}(q) = 1 + q_s/q$ in the limit of large q, where the boundaries have little influence. If the depletion layer is very thick, and all regions have the same dielectric constant, the presence of the gate modifies the usual result to

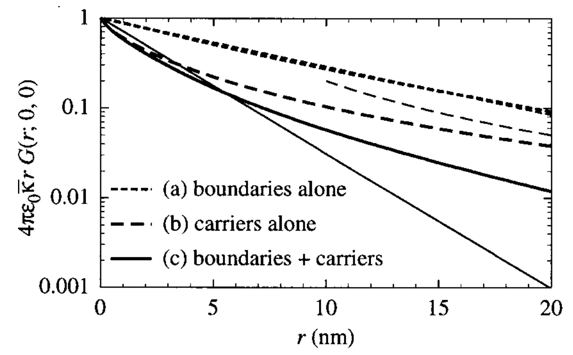

FIG. 2. Green's function $G(r;0,0)$ in real space for both points within the channel showing effect of screening by (a) boundaries alone, (b) carriers alone, and (c) both boundaries and carriers. The plot shows the ratio of $G(r;0,0)$ to its value in the dielectric media alone, $1/4\pi\epsilon_0\bar{\kappa}r$, which eliminates the $1/r$ divergence. Thick lines are the exact results while thin lines show the approximate expressions from Eqs. (2.7), (2.9), and (2.16).

$$\epsilon_{\mathrm{eff}}(q) = 1 + \frac{q_s}{q}[1-\exp(-2qd)], \tag{2.13}$$

which has been used in the theory of III–V heterostructures.[25]

A combination of Eqs. (2.4) and (2.11) gives the full Green's function for both points in the channel ($z=z'=0$),

$$\begin{aligned}[\tilde{G}(q;0,0)]^{-1} &= [\tilde{G}_0(q;0,0)]^{-1} + C_i \\ &= \epsilon_0 q(\kappa_{\mathrm{ox}}\coth qd + \kappa_{\mathrm{sc}}\coth qw) \\ &\quad + 2\epsilon_0\bar{\kappa}q_s.\end{aligned} \tag{2.14}$$

The transform to real space is plotted as curve (c) in Fig. 2. It diverges as $1/r$ for small distances, followed by a region where screening increases the decay to $1/r^3$, and at large distance the boundaries induce an exponential decay.

The Green's function can be combined with the approximation given in Eq. (2.5) to give

$$\begin{aligned}[\tilde{G}(q;0,0)]^{-1} &\approx 2\epsilon_0\bar{\kappa}(q_s + \sqrt{Q_s^2+q^2}) \\ &= C_i + \sqrt{(C_{\mathrm{ox}}+C_{\mathrm{sc}})^2 + (2\epsilon_0\bar{\kappa}q)^2}.\end{aligned} \tag{2.15}$$

Unfortunately this expression is awkward and it is much easier to use the inferior approximation

$$\begin{aligned}[\tilde{G}(q;0,0)]^{-1} &\approx 2\epsilon_0\bar{\kappa}\sqrt{(q_s+Q_s)^2+q^2} \\ &= \sqrt{(C_i + C_{\mathrm{ox}}+C_{\mathrm{sc}})^2 + (2\epsilon_0\bar{\kappa}q)^2}.\end{aligned} \tag{2.16}$$

The transform of this approximation to real space gives the same exponential decay as Eq. (2.7) but with $q_s + Q_s$ instead of Q_s alone. The error associated with this is shown by curves (c) in Fig. 2; the potential is too large at small distance, where it has its greatest effect, and decays too rapidly at large distance. This error proves unacceptable.

III. VARIANCE OF THE RANDOM POTENTIAL

The autocorrelation function of the potential $\phi(\mathbf{r},z)$ due to the random charges in the depletion layer is defined by

$$W(\mathbf{r};z,z') = \langle \delta\phi(\mathbf{r}',z)\,\delta\phi(\mathbf{r}'+\mathbf{r},z')\rangle, \tag{3.1}$$

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

where $\delta\phi$ is the deviation of ϕ from its mean value. This shows how the potential at any point $\mathbf{r}'$ is related to that at a point displaced by $\mathbf{r}$. It can be written in terms of the Green's function for the potential as

$$W(\mathbf{r};z,z') = e^2 \left\langle \sum_{i,j} G(\mathbf{r}'-\mathbf{r}_i;z,z_i) G(\mathbf{r}'+\mathbf{r}-\mathbf{r}_j;z',z_j) \right\rangle, \tag{3.2}$$

where i and j label the ionized impurities, each of charge $\pm e$ and situated at $(\mathbf{r}_i, z_i)$. This expression must be averaged over all space $(\mathbf{r}')$ and over all configurations of impurities, which is equivalent to an average over different devices. We assume that the impurities are distributed at random with uniform density $N(z)$ in each plane with a particular value of z. The average density $N(z)$ per unit volume varies with depth. After averaging, the Fourier transform of the correlation function within the channel $(z=z'=0)$ becomes

$$\tilde{W}(q) \equiv \tilde{W}(q;0,0) = e^2 \int_0^w |\tilde{G}(q;0,z)|^2 N(z)\, dz. \tag{3.3}$$

The variance is given by the correlation function at $r=0$,

$$\sigma^2 = W(0) = e^2 \int \frac{d^2\mathbf{q}}{(2\pi)^2} \int_0^w dz\, N(z) |\tilde{G}(q;0,z)|^2. \tag{3.4}$$

This can be rewritten using the relation (2.3) between $\tilde{G}(q;0,z)$ and $\tilde{G}(q;0,0)$ as

$$\sigma^2 = \frac{e^2}{2\pi} \int_0^\infty q\, dq |\tilde{G}(q;0,0)|^2 \int_0^w dz\, N(z) \frac{\sinh^2 q(w-z)}{\sinh^2 qw}. \tag{3.5}$$

The integral over z can often be performed analytically. Uniform doping of N_A throughout the depletion region gives

$$\sigma^2 = \frac{e^2 N_A}{4\pi} \int_0^\infty \frac{\sinh 2qw - 2qw}{q(\cosh 2qw - 1)} |\tilde{G}(q;0,0)|^2 q\, dq. \tag{3.6}$$

The final integration over q must be evaluated numerically if the exact Green's function from Eq. (2.4) or (2.14) is used but we shall now develop some analytic approximations.

A. Boundaries alone

Consider first a layer of δ doping with concentration $N_A^{(2D)}(z)$ at depth z and screening from the boundaries alone. This shows how different layers of donors contribute to the fluctuations. For small z, where the largest contribution arises, it is a good approximation to replace the hyperbolic functions in Eq. (3.5) with

$$\frac{\sinh^2 q(w-z)}{\sinh^2 qw} \approx e^{-2qz}. \tag{3.7}$$

Using Brew's approximation (2.5) for the Green's function, this gives a contribution to the variance of

$$\sigma^2(z) \approx \frac{N_A^{(2D)}(z)}{2\pi} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \int_0^\infty \frac{e^{-2qz} q\, dq}{Q_s^2 + q^2}$$

$$= \frac{N_A^{(2D)}(z)}{2\pi} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 g(2Q_s z). \tag{3.8}$$

Here, $g(x)$ is an auxiliary function associated with the sine and cosine integrals.[26] The variance diverges logarithmically as $z \to 0$, as noted by Brews[12] in his study of random charges at the silicon–oxide interface. Thus random charges close to the channel have the largest effect on the variance as expected, although the divergence would be suppressed if the finite thickness of the channel were included.

In the opposite limit z approaches w, the edge of the depletion layer. Straightforward expansion of Eq. (3.5) shows that $\sigma^2(z)$ vanishes like $(w-z)^2$. The contribution of these charges is suppressed by their images in the neutral region below the depletion layer, which we treat as a perfect conductor.

Now, consider a uniformly doped depletion layer. Eq. (3.6) for the variance with Eq. (2.5) for the Green's function yields

$$\sigma^2 = \frac{N_A}{4\pi} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \int_0^\infty \frac{\sinh 2qw - 2qw}{q(\cosh 2qw - 1)} \frac{q\, dq}{Q_s^2 + q^2}. \tag{3.9}$$

The quotient of hyperbolic functions can again be approximated by an exponential expression, which is chosen to have the same integral over q. This gives

$$\sigma^2 \approx \frac{N_A}{4\pi} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \int_0^\infty \frac{(1-e^{-qw})\, dq}{Q_s^2 + q^2} \tag{3.10}$$

$$= \frac{N_A}{4\pi Q_s} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \left[\frac{\pi}{2} - f(Q_s w)\right]. \tag{3.11}$$

Here $f(x)$ is another auxiliary function associated with the sine and cosine integrals. Its limiting values are $f(0) = \pi/2$ and $f(x) \sim 1/x$ as $x \to \infty$. This gives $\sigma = 71$ mV for the example, in good agreement with 69 mV from numerical evaluation of Eq. (3.6) with the full Green's function of Eq. (2.4).

Another approach, in the spirit of Brews' approximation to the Green's function, leads to a power spectrum that can be transformed simply to get the correlation function for a well-scaled device. The algebraic approximation

$$\frac{\sinh 2qw - 2qw}{q(\cosh 2qw - 1)} \approx \frac{1}{[(3/2w)^2 + q^2]^{1/2}} \tag{3.12}$$

is good for both small and large q. The variance is then

$$\sigma^2 \approx \frac{N_A}{4\pi} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \int_0^\infty \frac{q\, dq}{[(3/2w)^2 + q^2]^{1/2} (Q_S^2 + q^2)}. \tag{3.13}$$

Further, the constants in the denominator are close in value for a well-scaled device. This requires $(2/3) Q_s w \approx 1$; it is 3/2 for the example used here. The two terms can then be combined to give a trivial integral,

$$\sigma^2 \approx \frac{N_A}{4\pi} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \int_0^\infty \frac{q\, dq}{(\alpha^2 + q^2)^{3/2}} \tag{3.14}$$

$$= \frac{N_A}{4\pi Q_s} \left(\frac{e}{2\epsilon_0 \bar{\kappa}}\right)^2 \left(\frac{2}{3} Q_s w\right)^{1/3}, \tag{3.15}$$

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

where $\alpha^3=3Q_s^2/2w$. This gives $\sigma=70$ mV for the example, which is again in excellent agreement with 69 mV from numerical evaluation.

Finally, the asymptotic expansion for the limit $Q_s w \gg 1$, which holds for a very thin oxide, is

$$\sigma^2=\frac{N_A}{4\pi Q_s}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\left[\frac{\pi}{2}-\frac{1}{Q_s w}+\frac{\zeta(3)}{(Q_s w)^3}\cdots\right], \tag{3.16}$$

where $\zeta(3)=1.202$. This expansion confirms that most of the contribution to σ^2 comes from impurities close to the channel; the leading term alone gives an estimate for the example that rises only to $\sigma=81$ mV.

Another system of interest has an thin, undoped, epitaxial layer of thickness s in which the channel resides, with uniform doping of N_A in the layer $s<z<w$. The effect of the impurities removed from the epitaxial layer can be evaluated using Eq. (3.8) and subtracted from the result for uniform doping throughout the depletion layer. The difference is given by

$$\Delta\sigma^2(s)=\frac{N_A}{2\pi}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\int_0^s g(2Q_s z)\,dz$$
$$=\frac{N_A}{4\pi Q_s}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\left[\frac{\pi}{2}-f(2Q_s s)\right], \tag{3.17}$$

where $f(x)$ is the same auxiliary function as in Eq. (3.11). Even a thin epitaxial layer has a dramatic effect on the strength of fluctuations. For example, a thickness of only 2 nm reduces σ from 70 to 43 mV according to Eq. (3.17); numerical evaluation of Eq. (3.6) gives 44 mV. Further results will be given in Sec. V.

B. Carriers alone

We next consider the effect of screening by carriers alone. The Green's function is given by Eq. (2.8), and Eq. (3.5) for the variance becomes

$$\sigma^2=\frac{1}{2\pi}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\int_0^\infty\frac{q\,dq}{(q_s+q)^2}$$
$$\times\int_0^w dz\,N(z)\frac{\sinh^2 q(w-z)}{\sinh^2 qw}. \tag{3.18}$$

Again, we find that the contribution diverges logarithmically as $z\to 0$ and vanishes quadratically as $z\to w$.

We approximate the hyperbolic functions for a uniform layer of doping with the same exponential form as in Eq. (3.10). This leads to

$$\sigma^2\approx\frac{N_A}{4\pi}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\int_0^\infty\frac{(1-e^{-qw})dq}{(q_s+q)^2}$$
$$=\frac{N_A}{4\pi q_s}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2 h(q_s w), \tag{3.19}$$

where $h(x)=1-xe^xE_1(x)$ and $E_1(x)$ is the exponential integral.[26] Useful limits are $h(0)=1$ and $h(x)\sim 1/x$ for large x. For our example $q_s w=4.0$ and $\sigma=44$ mV; numerical evaluation of Eqs. (2.8) and (3.6) gives 43 mV. Thus, screening by carriers does not have a dramatic effect on the fluctuations at room temperature.

The algebraic approximation (3.12) to the hyperbolic functions in the previous section can again be used but the results are less compact. The asymptotic expansion for large $q_s w$ is

$$\sigma^2\sim\frac{N_A}{4\pi q_s}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2$$
$$\times\left[1-\frac{1}{q_s w}+\frac{\pi^2}{6(q_s w)^2}-\frac{3\zeta(3)}{(q_s w)^3}\cdots\right]. \tag{3.20}$$

The leading term alone gives $\sigma=49$ mV.

C. Boundaries and carriers

We have not found a satisfactory analytic approximation to the variance when screening by both boundaries and carriers is included. Numerical evaluation of Eq. (3.6) with the full Green's function (Eq. 2.14) gives $\sigma=40$ mV, so there is only a small reduction beyond 43 mV with screening alone. Unfortunately, the approximate Green's function in Eq. (2.16) leads to $\sigma=45$ mV, which is higher than the result with screening alone and is therefore unacceptable.

IV. AUTOCORRELATION FUNCTION OF THE RANDOM POTENTIAL

The autocorrelation function in real space is obtained by a Fourier transform of the power spectrum, whose integral was evaluated in the previous section to give the variance. This can rarely be done analytically. A convenient exception is the power spectrum in Eq. (3.14), derived for a channel with boundaries alone,

$$\tilde{W}(q)=\frac{N_A}{2}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\frac{1}{(\alpha^2+q^2)^{3/2}}. \tag{4.1}$$

Rotational symmetry in the $x-y$ plane leads to the usual Hankel transform,

$$W(r)=\frac{1}{2\pi}\int_0^\infty\tilde{W}(q)J_0(qr)q\,dq, \tag{4.2}$$

which gives

$$W(r)=\frac{N_A}{4\pi Q_s}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2(\tfrac{2}{3}Q_s w)^{1/3}\exp(-\alpha r). \tag{4.3}$$

The variance is given by $W(0)$ in agreement with Eq. (3.15). Thus, the autocorrelation function decays exponentially with a length scale $\alpha^{-1}=9.2$ nm for the usual example. This is in excellent agreement with numerical evaluation of the correlation function, shown in Fig. 3 where the two curves (a) are indistinguishable on this scale.

The correlation function decays asymptotically as a power law with screening by carriers alone (no boundaries), like the Green's function; the leading term is

$$W(r)\sim\frac{N_A}{4\pi q_s}\left(\frac{e}{2\epsilon_0\bar{\kappa}}\right)^2\frac{4q_s w}{3(q_s r)^2}. \tag{4.4}$$

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

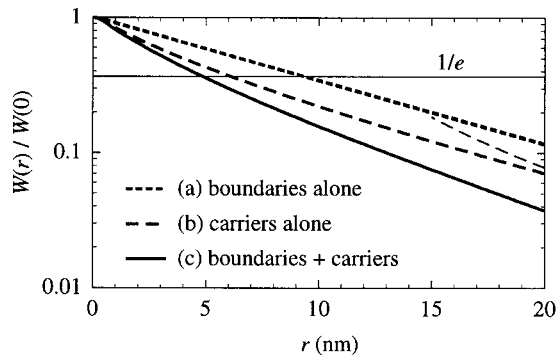

FIG. 3. Normalized correlation function $W(r)/W(0)$ in real space for both points within the channel showing effect of (a) boundaries alone, (b) carriers alone, and (c) both boundaries and carriers. Thick lines are the exact results while thin lines show the approximate expressions from Eqs. (4.3) and (4.4).

This is also shown as curve (b) in Fig. 3 but the asymptotic form is accurate only for larger r. The most rapid decay occurs when both screening and boundaries are included but we have found no simple expression for this.

The most important feature of the autocorrelation function is its characteristic length L because this sets the length scale of the fluctuations. A convenient definition, given the roughly exponential decay, is

$$L^2=\frac{1}{6}\frac{\int r^2 W(r)d^2\mathbf{r}}{\int W(r)d^2\mathbf{r}}=-\frac{1}{6}\frac{\nabla_q^2\tilde{W}}{W}\bigg|_{q=0}, \tag{4.5}$$

where the factor of $\frac{1}{6}$ ensures that $L=1/\alpha$ if $W(r)\propto\exp(-\alpha r)$. The result for a system with a uniformly doped layer, including screening by both boundaries and carriers, based on the full Green's function of Eq. (2.14), is

$$L^2=\frac{2}{9}\left[\frac{\kappa_{\rm ox}d+\kappa_{\rm sc}w}{\bar{\kappa}(q_s+Q_s)}+\frac{2w^2}{5}\right]. \tag{4.6}$$

This gives $L=9.0$ nm for boundaries alone and 6.9 nm for boundaries and carriers. These are in reasonable agreement with the radii at which the correlation functions plotted in Fig. 3 have decayed by a factor of $1/e$; Eq. (4.6) is less satisfactory when carriers are present because of the nonexponential decay.

Figure 4 shows the correlation length according to Eq. (4.6) as a function of doping, which changes the thickness of the depletion region and therefore Q_s. It is plotted in two ways. The thin lines are for a constant thickness of oxide, $d=3$ nm, while the thick lines are for a device scaled so that the thickness of the oxide is a constant fraction of the depletion thickness, $d=w/6$. For the scaled device with boundaries alone the length obeys $L\approx 1/\alpha\approx w/2$ and the corresponding curve therefore lies virtually on top of that for the depletion thickness. The two curves that include carriers are also nearly coincident, showing that the additional screening almost eliminates the effect of oxide thickness on the length scale of the fluctuations.

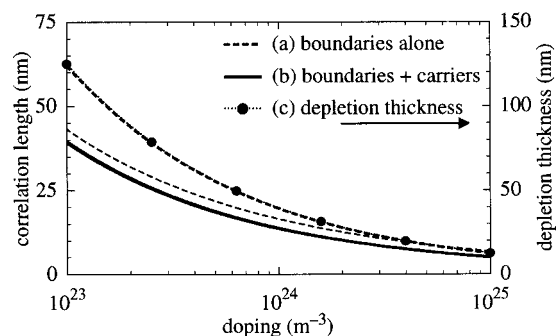

FIG. 4. Correlation length of potential fluctuations for (a) boundaries alone and (b) both boundaries and carriers as a function of doping. Thin curves are for a constant oxide thickness of $d=3$ nm while thick curves are for the thickness scaled to $d=w/6$. (c) Depletion thickness, shown against right-hand side scale as dotted line with symbols.

V. COMPARISON WITH NUMERICAL SIMULATION

We have compared our results with three-dimensional atomistic simulations of MOSFETs.[27,28] In this approach, the impurities are treated as discrete charges, distributed at random with the appropriate average concentration. Quantities such as the threshold voltage can be determined for each device. The simulation must be repeated with a large number of configurations of random impurities to accumulate statistics for the mean value and the fluctuations between devices. This technique is time consuming and guidance from the analytical approach presented here is therefore valuable.

We must first address the relation between the random potential that we have studied here and the threshold voltage deduced from the simulations. One picture is to imagine that we have considered the random potential in a single device of infinite area. Individual, finite devices can then be made by "punching out" appropriate areas of this random potential. The relation between the potential in an individual device and our results depends on the size of the device compared with the correlation length.

(1) If individual devices are *small* compared with the correlation length, the potential is almost constant within each device. The variations between devices are then given directly by the statistics that we have calculated for an infinite area. For example, the standard deviation of the threshold voltage is given by Eq. (3.15) with no further corrections.
(2) It is harder to treat the opposite limit, where individual devices are *large* compared with the correlation length. The average potential within each device is smoothed out because there will be several peaks and valleys, reducing its standard deviation from Eq. (3.15) by a factor of roughly (correlation length)/(length of device).

A well-scaled device falls between these limits and we are not aware of any complete theory for relating threshold voltage to potential fluctuations under these conditions. For example, most of the simulations used a device with an effective area of $(50\text{ nm})^2$ and the parameters listed at the end of Sec. I, which give a correlation length $L\approx 10$ nm. At first sight, it appears that the device is large compared with L but the critical region near threshold is much smaller, a barrier about 20 nm long between source and drain. Fluctuations turn the barrier into an irregular mountain range and current

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

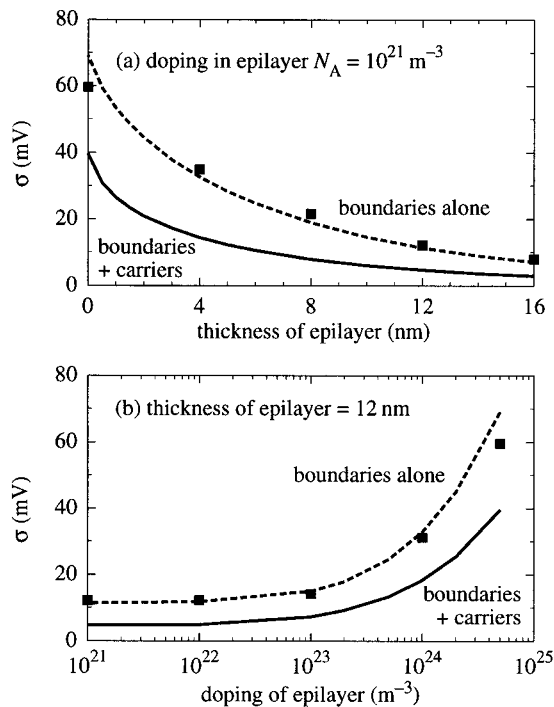

FIG. 5. Standard deviation of fluctuations as a function of (a) thickness and (b) doping of an epitaxial layer. Lines show the random potential from calculations in this article and symbols show the threshold voltage from atomistic simulations (see Ref. 27).

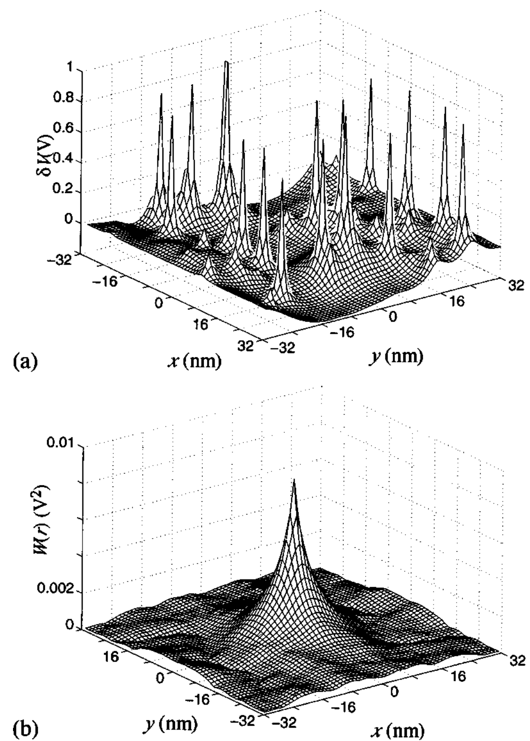

FIG. 6. (a) Typical example of the random potential near threshold at the oxide–semiconductor interface in an atomistic simulation and (b) its autocorrelation function.

starts to flow when the first pass through the mountains becomes accessible. In this case, the standard deviation of the random potential that we have calculated is itself a reasonable guide to that of the threshold voltage, and we shall compare them directly. It might be more accurate to reduce the standard deviation of the random potential by a factor of $[(2L)/(\text{width of device})]^{1/2}$ to allow for several passes through the mountain but we have not made this correction because of other difficulties in the comparison, described later in this section.

There are very few carriers at threshold and the results for boundaries alone are therefore appropriate, but for completeness we shall also show our results that include screening by $N_i = 10^{16}$ m^{-2} carriers in the channel.

We first consider the variation of threshold voltage with density of doping, N_A. This was fitted to a power law in the simulations,[27] $\sigma \propto N_A^p$, and the best fit was $p = 0.40$. The general result for the variance of the random potential, Eq. (3.6), would give $p = \frac{1}{2}$ from the prefactor of N_A if everything else remained constant. However, an increase in doping reduces the thickness w of the depletion layer, which in turn increases the screening wave vector Q_s. These lessen the increase in σ and we find by fitting our calculations that the power is reduced to 0.47 in the practical range of doping. This is still higher than the power law for the threshold voltage deduced from the simulations. We suspect that the discrepancy arises from the change in correlation length, which gets shorter as the doping increases and gives more averaging.

It was mentioned in Sec. III A that fluctuations in the channel can be suppressed by growing a lightly doped epitaxial layer ($N_A = 10^{21}$ m^{-3}) on top of the highly doped substrate ($N_A = 5 \times 10^{24}$ m^{-3}). The effect of this as a function of thickness s is shown in Fig. 5(a). The standard deviation of the potential was obtained by numerical integration of Eq. (3.5), employing the full Green's function from Eq. (2.14). Agreement between these results and the simulations of the threshold voltage is remarkably good, and shows that an epitaxial layer of 4 nm thickness cuts the fluctuations by a factor of 2. The effect of the concentration of acceptors in an epitaxial layer of thickness $s = 12$ nm is shown in Fig. 5(b). Again, there is good agreement between the simulations of the threshold voltage and our calculations of the random potential. A concentration below 10^{23} m^{-3} in the epitaxial layer is sufficient to suppress the fluctuations.

We have also looked more closely at the random potential at the silicon–oxide interface in simulated devices. These had the same parameters that we have used with an area of $(64 \text{ nm})^2$ and the gate voltage was adjusted to be near threshold. Figure 6(a) shows a typical potential. There is a smooth variation along y from source to drain, on top of which are peaks from individual acceptors whose height depends on their distance from the interface. The standard deviation was found to be 94 mV, rather larger than our predicted value of 69 mV.

The correlation function of this random potential was calculated from ten samples. A discrete two-dimensional (2D) Fourier transform of the random potential fluctuation was taken over x and y and the correlation function was constructed by summing the squares of the Fourier components and averaging over the samples. Finally, a 2D discrete

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

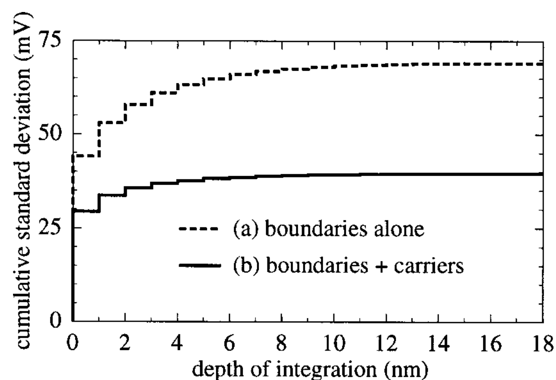

FIG. 7. Standard deviation of the random potential as a function of thickness included in the integration.

inverse Fourier transform was performed to obtain the correlation function in real space, which is plotted in Fig. 6(b). It follows the predicted exponential form well, but the correlation length is only about 6 nm rather than the 9 nm that we predict.

We believe that these discrepancies arise because the simulation is based on a mesh while the analytic results are for a continuum. Figure 7 shows how the standard deviation of the random potential builds up in the analytical approach as the region of integration is extended from the silicon–oxide interface in slabs of thickness 1 nm. More than half of the total standard deviation arises from the first 1 nm slab, which represents a single grid spacing in the simulations. This conclusion is supported by the potential plotted in Fig. 6(a). All the tall peaks have virtually the same height because they arise from the first layer of impurities; the second layer provides a set of much lower peaks, and subsequent layers can not be distinguished. Moreover, the first layer of impurities lies at the silicon–oxide interface and a continuous Coulomb potential would diverge at each impurity. The divergence is suppressed by the discrete approximation to Poisson's equation in the simulation, but these impurities contribute a short-ranged peak to the "typical" Coulomb potential. This raises both the standard deviation and the peak at the origin of $W(r)$, which makes the apparent correlation length shorter. Better agreement between simulations and the analytic approach would be expected for a finer grid, but a reduction of 2 would make the spacing equal to the lattice constant of silicon. It is then arguable that the analytic approach should itself be based on a sum over atomic sites rather than a continuum.

The dominant contribution of the impurities within 1 nm of the silicon–oxide interface also shows that the results are insensitive to assumption 5 in our model (Sec. I), that the bottom of the depletion region can be treated like a flat, conducting plate. Screening by the nearby gate is much more important.

VI. CONCLUSIONS

We have provided analytic estimates of the magnitude and length scale of the fluctuating potential in the channel of a MOSFET due to the random distribution of ionized impurities in the depletion region and shown how they can be related to fluctuations of the threshold voltage between devices. The results are in reasonable agreement with atomistic simulations but precise comparison is difficult because of the finite grid used in the simulations. A finer grid may be necessary for the most precise results, and similarly it may be necessary to base analytic work on a grid rather than a continuum. We find that a lightly doped, epitaxial layer reduces the fluctuations drastically, in good agreement with numerical studies.[27]

The calculations can be extended to match more recent simulations. It should be relatively straightforward to include screening in a polysilicon gate rather than the metal gate used in our model.[29] This is particularly important for a thin oxide. The thickness of the channel is far from negligible in modern MOSFETs and the potential should be averaged over the distribution of carriers rather than taken at the silicon–oxide interface alone.[30] Fluctuations in the quantum mechanical energy level due to variations in the normal electric field[31] could also be included. A major challenge remains an extension of the theories that deduce fluctuations in the threshold voltage from those in the potential to take full account of the length scale of the fluctuations compared with the dimensions of a well-scaled transistor.

ACKNOWLEDGMENT

This work was supported by NASA Ames Research Centre through Grant No. NAG 2-1241.

[1] B. Hoeneisen and C. A. Mead, Solid-State Electron. **15**, 819 (1972).
[2] R. W. Keyes, Proc. IEEE **63**, 740 (1975).
[3] T. Mizuno, J. Okamura, and A. Toriumi, IEEE Trans. Electron Devices **41**, 2216 (1994).
[4] M. Steyaert, J. Bastos, R. Roovers, P. Kinget, W. Sansen, B. Graindourse, A. Pergot, and E. Janssens, Electron. Lett. **30**, 1546 (1994).
[5] J. T. Horstmann, U. Hilleringmann, and K. F. Gosser, IEEE Trans. Electron Devices **45**, 299 (1997).
[6] K. Nishiohara, N. Shiguo, and T. Wada, IEEE Trans. Electron Devices **39**, 634 (1992).
[7] H.-S. Wong and Y. Taur, Tech. Dig. - Int. Electron Devices Meet. 705 (1993).
[8] P. A. Stolk, F. P. Widdershoven, and D. B. M. Klaassen, IEEE Trans. Electron Devices **45**, 1960 (1998).
[9] A. Asenov, IEEE Trans. Electron Devices **45**, 2505 (1998).
[10] A. Asenov, G. Slavcheva, A. R. Brown, R. Balasubramaniam, and J. H. Davies, Superlattices Microstruct. **27**, 215 (2000).
[11] E. H. Nicollian and A. Goetzberger, Bell Syst. Tech. J. **46**, 1055 (1967).
[12] J. R. Brews, J. Appl. Phys. **43**, 2306 (1972).
[13] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology* (Wiley, New York, 1982).
[14] V. A. Gergel' and R. A. Suris, Sov. Phys. JETP **48**, 95 (1978).
[15] V. A. Gergel' and R. A. Suris, Sov. Phys. JETP **57**, 415 (1983).
[16] G. Slavcheva and I. Yanchev, Solid State Commun. **79**, 439 (1991).
[17] J. R. Brews, J. Appl. Phys. **46**, 2181 (1975).
[18] J. R. Brews, J. Appl. Phys. **46**, 2193 (1975).
[19] A. O. Orlov, A. K. Savchenko, and B. I. Shklovskii, Sov. Phys. Semicond. **23**, 830 (1989).
[20] J. H. Davies and G. Timp, *Heterostructures and Quantum Devices*, edited by N. G. Einspruch, and W. R. Frensley (Academic, San Diego, 1994), p. 385.
[21] K. Takeuchi, 1998 Symposium on VLSI Technology, Honolulu, 9–11 June 1998, p. 72
[22] F. Stern, Phys. Rev. Lett. **18**, 546 (1967).
[23] F. Stern and W. E. Howard, Phys. Rev. **163**, 816 (1967).

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

[24] See, for example, J. M. Ziman, *Elements of Advanced Quantum Theory* (Cambridge University Press, Cambridge, UK, 1970).
[25] J. H. Davies and I. A. Larkin, Phys. Rev. B **49**, 4800 (1994).
[26] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1970), Chap. 5.
[27] A. Asenov and S. Saini, IEEE Trans. Electron Devices **46**, 1718 (1999).
[28] A. Asenov, A. R. Brown, J. H. Davies, and S. Saini, IEEE Trans. Comput.-Aided Des. **18**, 1558 (1999).
[29] A. Asenov and S. Saini, IEEE Trans. Electron Devices **47**, 805 (2000).
[30] A. Asenov, G. Slavcheva, A. R. Brown, J. H. Davies, and S. Saini, IEEE Trans. Electron Devices **48**, 722 (2001).
[31] T. H. Ning and C. T. Sah, Phys. Rev. B **9**, 527 (1974).

Downloaded 09 Jan 2007 to 131.227.76.18. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp