Study of direct tunneling through ultrathin gate oxide of field effect transistors using Monte Carlo simulation

E. Cassan, S. Galdin, P. Dollfus, and P. Hesto
Institut d’Electronique Fondamentale, UMR CNRS 8622, Université Paris Sud, Bât. 220 91405 Orsay cedex, France

(Received 22 February 1999; accepted for publication 22 June 1999)

Direct tunneling gate currents of ultrathin gate oxide thickness metal oxide semiconductor field effect transistors (MOSFETs) are modeled in a two-step calculation procedure based on the treatment of physical microscopic data acquired during Monte Carlo device simulation. Gate currents are obtained by weighting the carrier perpendicular energy distribution at the Si/SiO$_2$ and N$^+$-poly–Si/SiO$_2$ interfaces by the electron transmission probability, which is calculated by the one-dimensional Schrödinger equation resolution with the transfer-matrix method. The procedure is applied to a 0.07 $\mu$m gate length and 1.5 nm gate oxide thickness transistor, for which the gate and drain voltage influences on gate currents are studied by assuming at first a uniform gate oxide layer. It is shown that the maximum gate current is obtained for one of the two static points of complementary metal oxide semiconductor inverters: $V_{GS}=V_{DD}$ and $V_{DS}=0$, which raises a severe problem of standby power consumption. The contribution of hot carriers to the tunnel current is evaluated and is found to be small in case of such ultrathin oxide $n$-MOSFETs: contrary to thick (>5 nm) gate oxide transistors, the maximum gate current is not linked to the carrier energy peak in the channel but is located near the source well where the electron concentration is the largest. Oxide thickness fluctuations are then considered by meshing the oxide surface area and assuming a Gaussian law for the local oxide thickness deviation to the mean value. It is shown that a correct agreement is achieved with experimental published data when the oxide film nonuniformity is included in the calculation. Gate currents mapping for different bias conditions are given and analyzed, which show that very high current densities run through the oxide layer in the vicinity of weak points. An estimate of the surface through which flows the major part of the current is made, and a link between the highly nonuniform current leakage and the soft-breakdown mechanism of the oxide layer is proposed. © 1999 American Institute of Physics. [S0021-8979(99)01519-4]

I. INTRODUCTION

With the downsizing of the metal oxide semiconductor field effect transistor (MOSFET), the gate oxide thickness has continued to decrease in order to achieve a good control of the channel charge and to provide against short channel effects. For many years this decrease has been applied without special care to the gate current because the oxide was thick enough to prevent a large flow of charges through it, under normal operating conditions (no active stress). Yet, in manufacturing, oxide thicknesses of 5 nm are now customary and research currently deals with oxide in the range of 1.5–3 nm. Below 3 nm, the potential barrier between the silicon substrate and the gate material cannot be considered anymore as impenetrable and a direct tunneling current through it becomes the main leakage mechanism. 1–8 In spite of this, it has been proposed to investigate the scaling of the gate oxide thickness of MOSFETs below 3 nm. 1–3 Momose et al. 1–3 have proved that the conventionally perceived limitations on oxide thickness should not be considered as an absolute limitation. Indeed they have shown that good device performance can be achieved with 1.5 nm gate oxide $n$-MOSFETs with a supply of about 1 V at room temperature, in spite of a gate current density of some tens of A/cm$^2$. They have obtained drive currents of 1200 $\mu$A/$\mu$m and transconductance of 1000 $\mu$S/$\mu$m for such devices. These results can be attributed to the short gate length scaled with such an ultrathin gate oxide: 0.1 $\mu$m. Indeed, for a constant value of the oxide thickness, decreasing the gate length limits the total tunneling current and simultaneously leads to very high drive currents, thus improving the drive current over gate current ratio: typically 10$^5$. The same authors have examined items linked to power consumption and to oxide reliability: 1–3 they have studied the nonuniformity of an oxide film of 1.5 nm grown by rapid thermal annealing. All these interesting experimental results explain why a quantitative understanding of the tunneling phenomena is an important issue in design and optimization of future $n$-MOSFETs.

The general feature of gate tunnel current modeling is to couple the current calculation to a device simulator, whose role is to deliver all the physical microscopic data necessary to the local tunnel charge flow evaluation. Recently, this procedure has been proposed for the study of gate currents of MOS capacitors and MOSFET transistors. 8 Yet, no results about very short MOSFETs have been given in this article, where a drift-diffusion transport model is the basis of the
The aim of this article is to calculate gate currents of ultrashort MOSFETs with ultrathin gate oxides for different gate and drain voltages, to study their lateral distributions from source to drain and their local densities, whose evaluation is important for reliability concerns. We are particularly interested in the carrier distribution function at the semiconductor/oxide interface because we want to clarify the role of hot carriers in gate currents of ultrashort MOSFETs. For this reason a Monte Carlo (MC) device simulator has been chosen, because it allows an accurate description of the carrier energy: especially, ballistic or near-ballistic carriers are naturally taken into account. Finally, we would like to clarify the effect of the oxide film nonuniformity on gate currents.

This article is organized as follows. In Sec. II the model of gate current calculation is presented: gate currents are calculated thanks to microscopic data acquired in a first step during MC device simulation. The validity and precision of all the physical hypotheses made to evaluate the transmission probability of each carrier hitting the interface are discussed. The procedure applied to acquire the relevant physical data is presented, and the bulk-to-metal and metal-to-bulk gate current expressions used to evaluate the current leakage through the oxide are given. In Sec. III the studied n-MOSFET transistor is presented and is electrically characterized by a drain current over gate voltage characteristic for a constant drain voltage. Then, the gate voltage influence on gate current for a zero drain voltage is examined, in order to evaluate the gate current order of magnitude. The effect of the drain voltage is also studied and different hypotheses are developed to explain the obtained results. Finally, Sec. IV examines the question of the oxide film thickness nonuniformity: the model of oxide thickness variation is described in a first step, and results are given about the tunnel current spatial distribution over the gate oxide surface for different bias points.

II. MODEL OF GATE CURRENT CALCULATIONS

In this section we explain why the approach of the transmission probability is adopted and we develop the different hypotheses made to evaluate it. No analytical expression of this physical quantity is given because it has been numerically calculated by the resolution of the one-dimensional Schrödinger equation. The calculation procedure of the gate current is explained and the two expressions used for bulk-to-metal and metal-to-bulk gate current evaluation are given.

A. Transmission probability

In MOSFET transistors, carriers are confined in the narrow potential well close to an inverted silicon surface. Tunneling physically corresponds to the penetration of carrier wave functions into the oxide and to their tails into the metal region. That is the reason why any calculation of the tunnel current should be theoretically made by properly solving this quantum-mechanical problem, which necessitates the self-consistent resolution of the Poisson and Schrödinger equations. Several publications deal with this task for MOS capacitors,5–9 yet it is far more difficult in the case of MOSFET transistors. Moreover, calculations made for capacitors with quantum-mechanical models6–8 have given very similar currents to the ones based on conventional models.4 We have thus decided to adopt the approach of the transmission probability. Carriers are considered as classical particles in the semiconductor channel, but the transmission probability, which is a wave-mechanic phenomenon, describes their ability to cross the tunnel barrier when they impinge on it.

The physical ground of the transmission probability calculation is to assimilate the incident carrier to a planar wave, and to calculate the attenuation of its wave function by solving the one-dimensional Schrödinger equation in the normal direction (i.e., parallel to the tunnel current).10,11 Practically, the tunnel barrier profile is approximated by a piecewise-constant potential barrier11 and the one-dimensional Schrödinger equation is solved by properly matching the wave function and the probability density current each time the potential energy varies: a two-dimensional matrix can then be written and numerically calculated, from which the transmission probability can be deduced. The wave attenuation strongly depends on the carrier quantities conserved during tunneling, which have thus to be clearly specified: in particular, is the parallel wave vector conserved or not? This question is difficult because it deals with properties of the intermediate layer, i.e., SiO2, whose band structure is not very well known, especially for very short thickness. It has been addressed by Weinberg in a series of papers,12–14 who has reported experiments showing that kL is relaxed in case of metal/SiO2/Si structures. We have thus applied the conservation of the total and perpendicular carrier energies during tunneling, like in Ref. 4. This approach makes the transmission problem truly one dimensional, and the transmission probability only depends on the normal carrier energy, as in the case of the WKB approximation.

The two crucial material parameters are the barrier height $\Phi_B$ and the tunneling mass $m^*_{ox}$. On the basis of Weinberg works,12–14 we have chosen $\Phi_B = 3.1$ eV and $m^*_{ox} = 0.5 \, m_0$ ($m_0$ being the free electron mass), which are the most frequently used values in literature (see, for example, Refs. 4, 5, 15, 16–20). Finally, an approach must be defined about the image force. It has been concluded in Refs. 4, 21, and 16 that the classical image potential could be used after having removed the singularities which arise at the edges of the barrier. On the contrary, it has been shown that no image potential correction is more consistent.12–14 In fact, the question of the image potential correction is still controversial. From our point of view, there is too much uncertainty about the band structure of the oxide layer to apply the classical image potential correction to the barrier of 3.1 eV. We think that the parameters characterizing the oxide layer should be considered as a whole set of parameters. Because the values of $\Phi_B = 3.1$ eV and $m^*_{ox} = 0.5 \, m_0$ have been determined without the image correction,12 we think that no image correction has to be applied to these values. Thus we have not considered the effect of image force in our calculations.

Figure 1 shows the potential barrier seen by electrons for
a positive gate voltage in case of a Si/SiO\textsubscript{2}/N\textsuperscript{+}-polysilicon capacitor. A silicon mass of 0.258 m\textsubscript{0} (conduction mass) and a polysilicon mass of m\textsubscript{0} have been assumed for the transmission probability evaluation. Indeed, some trials have confirmed to us the very small influence of these two parameters on the transmission probability, as noticed in Ref. 4. We see that, all material parameters being specified, the probability only depends on the oxide thickness t\textsubscript{ox}, on the oxide field F\textsubscript{ox}, and naturally on the perpendicular energy E\textsubscript{\perp} of the incident carrier.

**B. Physical data acquisition**

A device simulator is needed to describe all the physical microscopic data necessary to gate current calculations. Rigorously, the calculation of the gate current must be implemented into the transport model, which has to take the gate leakage as an additional physical condition. A major simplification can be made if we notice that gate currents of ultrashort MOSFETs are weak enough to constitute a very small correction to the device operating conditions for any bias point. Indeed, a gate current of 1.6 A/cm\textsuperscript{2}, which is the order of magnitude of tunneling currents for such transistors,

a silicon mass of 0.258 m\textsubscript{0} (conduction mass) and a polysilicon mass of m\textsubscript{0} have been assumed for the transmission probability evaluation. Indeed, some trials have confirmed to us the very small influence of these two parameters on the transmission probability, as noticed in Ref. 4. We see that, all material parameters being specified, the probability only depends on the oxide thickness t\textsubscript{ox}, on the oxide field F\textsubscript{ox}, and naturally on the perpendicular energy E\textsubscript{\perp} of the incident carrier.

**B. Physical data acquisition**

A device simulator is needed to describe all the physical microscopic data necessary to gate current calculations. Rigorously, the calculation of the gate current must be implemented into the transport model, which has to take the gate leakage as an additional physical condition. A major simplification can be made if we notice that gate currents of ultrashort MOSFETs are weak enough to constitute a very small correction to the device operating conditions for any bias point. Indeed, a gate current of 1.6 A/cm\textsuperscript{2}, which is the order of magnitude of tunneling currents for such transistors,

Practically, the channel length is divided into n\textsubscript{L} Si/SiO\textsubscript{2} interface areas and the energy range is also numerically sampled with an energy step ΔE\textsubscript{L}. The oxide electric field and the number N\textsubscript{△T,i\textsubscript{L}}(E\textsubscript{i\textsubscript{L}}) of carriers hitting the Si/SiO\textsubscript{2} interface during ΔT with a perpendicular energy (i.e., the kinetic energy in the direction of the tunneling current) between E\textsubscript{i\textsubscript{L}} and E\textsubscript{i\textsubscript{L}}+ΔE\textsubscript{L} are acquired for each interface area i\textsubscript{L}. The flow of hitting carriers at the i\textsubscript{L}\textsuperscript{th} interface area and having a perpendicular energy between E\textsubscript{i\textsubscript{L}} and E\textsubscript{i\textsubscript{L}}+ΔE\textsubscript{L} is computed as Φ\textsubscript{i\textsubscript{L}}(E\textsubscript{i\textsubscript{L}})=[N\textsubscript{△T,i\textsubscript{L}}(E\textsubscript{i\textsubscript{L}})/S\textsubscript{i\textsubscript{L}} ΔT], where S\textsubscript{i\textsubscript{L}} is the surface of the i\textsubscript{L}\textsuperscript{th} area. Notice that Φ\textsubscript{i\textsubscript{L}}(E\textsubscript{i\textsubscript{L}}) is independent of ΔT. Then, the bulk to metal gate current of each area is calculated by weighting the flow of carriers by the transmission probability P, according to the following expression where q is the unitary charge value:

$$J_{i\textsubscript{L}}^{\text{SC→metal}}=q \sum_{j} \Phi_{i\textsubscript{L}}(E_{j\textsubscript{L}}) P_{i\textsubscript{L}}(E_{j\textsubscript{L}}), \quad 1 \leq i\textsubscript{L} \leq n\textsubscript{L}. \quad (1)$$

The metal to gate current is expressed in the same way, except that an analytic carrier energy distribution is assumed in the metal: the Fermi one. The summation over perpendicular energy is then continuous:

$$J_{i\textsubscript{L}}^{\text{metal→SC}}=\frac{q k_B T_{\text{lattice}} m_{\text{metal}}}{2 \pi^2 h^3} \int P_{i\textsubscript{L}}(E_{\perp})$$

\[ \exp \left( \frac{(E_F-E_C) m_{\text{metal}} E_{\perp}}{k_B T_{\text{lattice}}} \right) \] \[ dE_{\perp}. \quad (2) \]

**III. RESULTS FOR A CONSTANT OXIDE THICKNESS**

In this section, the ultrathin gate oxide transistor chosen is described and is electrically characterized by a drain current over gate voltage characteristics for a supply drain voltage of 1 V. Then, the influence of both gate and drain voltages on gate currents are described.
Gate currents are calculated for an $n$-MOSFET transistor of 0.07 $\mu$m effective channel length (the SiO$_2$ layer overlaps the source and drain wells of 2 nm each side), 1.5 nm gate oxide thickness, and 30 nm junction depth (see Fig. 2). Uniform doping profiles are assumed for simplicity, either in source and drain wells ($5 \times 10^{19}$ cm$^{-3}$, $N$ type), as in bulk ($10^{18}$ cm$^{-3}$, $P$ type). These values have been suggested by recent experimental works. These high doping concentrations are typical of ultrashort MOSFETs, for which short channel effects have to be controlled. Figure 3 shows the drain current over gate voltage characteristics for a drain voltage of 1 V. About 80,000 particles were initially implanted in the device simulated in this work. A particle multiplication technique has been used to statistically enhance the number of rare events in the MOSFET channel at low gate voltage. Indeed, this technique is necessary to obtain a clear subthreshold characteristic by the MC method. We see in Fig. 3 that a drive current of about 1050 $\mu$A/\mu$m is achieved for the high supply voltage $V_{DD}$ of 1 V, which is near experimental values published for such a tunnel gate oxide transistor.

We now begin the gate current study by the gate voltage influence on tunneling for a constant drain voltage: $V_{DS}=0$ is chosen in order to evaluate only the gate voltage effect. Figure 4 shows the gate current absolute value over gate voltage characteristics for a zero drain voltage. The increase of gate current when the gate voltage is raised from zero to $V_{DD}$ can be explained as follows: because of the zero drain bias, the bulk electron population has a rather cold perpendicular energy distribution at the Si/SiO$_2$ interface for any gate voltage, but the higher and higher number of hitting carriers (deeper and deeper inversion) and the transmission probability rising with the oxide field growing are responsible for the bulk to metal gate current increase. The gate current variation when negative gate voltages are applied can be explained by the oxide field influence on the transmission probability, which is now favorable to injection from the metal electrode. Nevertheless, we notice a sharp difference between the evolution of gate currents obtained for negative and positive gate voltages. Indeed, when a gate voltage greater than approximately 0.1 V is applied to the gate electrode (small inversion), the potential under the oxide is almost constant all along the channel and equal to zero because of the zero drain bias. The voltage drop between the $N^+$ poly-Si electrode and the channel is thus approximately equal to the gate voltage all along the channel, which greatly favors the bulk to metal charge injection. On the contrary, when a negative gate voltage is applied, no inversion layer is formed under the oxide (space charge area). The bulk potential at the Si/SiO$_2$ interface even becomes negative in a large part of the channel, i.e., of the same polarity than the gate electrode, which lowers the oxide potential drop and thus the tunnel injection from metal. This explains the relatively low gate currents observed for negative gate voltages.

It can be concluded from the Fig. 4 analysis that gate currents of about 1 A/cm$^2$ (a few nA/\mu$m for a 0.07 $\mu$m gate length transistor) can be expected for a gate voltage of less than 1 V and for a gate oxide layer of 1.5 nm. These results appear to be about one order of magnitude smaller than values published in literature for comparable devices, but they will be updated in Sec. IV where technological fluctuations of the oxide layer thickness are taken into account.

Our purpose is now to understand the drain voltage influence on the gate current. Two rather different situations occur: the first corresponds to the $n$-MOSFET “off-state,” i.e., to a zero gate voltage, whereas the second one corresponds to the $n$-MOSFET “on-state,” i.e., to $V_{GS}=V_{DD}$. In the first case, the gate current is mainly from metal to bulk. As the electron population is in thermal equilibrium in the metal electrode, no carrier flow dependence at the metal/oxide interface with the drain voltage exists, and the tunneling current is thus simply modulated through the transmission probability dependence. When a positive drain voltage

![FIG. 2. $n$-MOSFET transistor studied: bulk Si ($P$-doped, $10^{18}$ cm$^{-3}$), source and drain contacts ($N$-doped, $5 \times 10^{19}$ cm$^{-3}$). Gate, source, and drain metallic contacts are not drawn. Moreover, Monte Carlo device simulation are only two dimensional and the transistor width is only introduced for the oxide film nonuniformity description in Sec. IV.](image1)

![FIG. 3. Drain current over gate voltage characteristic for $V_{DS}=V_{DD}=1$ V.](image2)

![FIG. 4. Gate current over gate voltage characteristic for $V_{DS}=0$.](image3)
is applied, the oxide electric field is weak except near the channel end where the silicon potential reaches the drain contact voltage. The metal to bulk tunneling injection is thus only locally favored. That is the reason why a quite small gate current is obtained in this case: $1.15 \times 10^{-2}$ A/cm$^2$ ($8 \times 10^{-3}$ nA/µm) for the studied transistor at $V_{GS}=0$ and $V_{DS}=V_{DD}=1$ V.

The drain voltage influence on the gate current is more complex in the second case ($V_{GS}=V_{DD}$). Indeed, the drain bias simultaneously modulates the electron energy distribution at the Si/SiO$_2$ interface and the potential barrier felt by electrons all along the channel. More precisely, the carrier population is heated by the longitudinal electric field due to the positive drain voltage, and this is favorable to a gate current increase. In the same time, the number of hitting carriers decrease from source to drain (decreasing electron concentration), and the potential barrier is less and less favorable to bulk to metal tunneling, as is illustrated in Fig. 5 for $V_{GS}=V_{DS}=V_{DD}$. Figure 6, which shows the transmission probability in three channel points of the previously described $n$-MOSFET, gives a quantitative evaluation of this barrier sharpness increase from source to drain. These last two effects tend on the contrary to lower the silicon to polysilicon gate current. The problem is knowing what is the preponderant tendency between the two.

For thick gate oxide layer transistors (more than 5 nm), the probability of carrier injection is extremely small in any point of the channel (even near source), except for very hot carriers. In this case, only carriers having an energy of at least 2 eV significantly contribute to the gate current, either by trap-assisted$^{19,20}$ or nonassisted tunneling phenomena (Fowler–Nordheim tunneling and thermionic emission$^{23–27}$). That is the reason why in this case, high energy tails are so important to gate current calculations. Electron injection is then closely related to the mean electron energy maximum in the channel, i.e., near drain.$^{24–26}$

The case of ultrathin gate oxide layer transistors (less than 3 nm) at small supply voltages, for which the direct tunneling current is the preponderant cause of gate leakage, is now studied on the basis of the previously described transistor. Figure 7 presents the gate current versus drain voltage characteristic for $V_{GS}=V_{DD}=1$ V (solid line), and clearly indicates a gate current decrease when the drain voltage is raised. This means that the hitting carrier flow lowering and the potential barrier effect along the channel are stronger than the carrier heating phenomena. Indeed, contrary to the case of thick oxide layers, electrons do not need to be very hot to be injected from Si into SiO$_2$ through an ultrathin layer ($<3$ nm). That is the reason why they can be injected near source where the electron concentration is the largest, whereas in the same time the barrier effect disadvantages the charge injection near the drain: see Fig. 6. Moreover, the small supply voltage does not allow hot carriers to acquire high perpendicular energies, as shown in Fig. 8, which presents the carrier perpendicular energy distribution at the Si/SiO$_2$ interface for $V_{GS}=V_{DS}=V_{DD}=1$ V and for the same positions into the channel than those of Fig. 6: the nearer the drain the farther is the population from equilibrium. Very near drain, a partition in two electron groups is particularly clear: those accelerated from source and responsible for the

FIG. 5. Schematic potential barrier along the channel for $V_{GS}=V_{DS}=V_{DD}$.

FIG. 7. Gate current vs drain voltage for $V_{GS}=V_{DD}=1$ V with a uniform (solid line) and with nonuniform (dashed line) gate oxide thickness.

FIG. 6. Transmission probability vs electron perpendicular energy in three channel points of the $n$-MOSFET transistor described in Sec. III at $V_{GS}=V_{DS}=1$ V: near source, in the channel middle, near drain.

FIG. 8. Perpendicular electron energy distribution at the Si/SiO$_2$ interface for $V_{GS}=V_{DS}=1$ V, near source, in the middle of the channel, and at the channel end.
curve tail around 0.2 eV, and those essentially diffused from drain in the channel. Yet, the carrier heating along the channel is clearly too small to counterbalance the transmission probability decrease given in Fig. 6. The consequence of the above arguments is that, contrary to what is true for thick gate oxide transistors, tunnel bulk-to-metal injection is stronger near source than near drain for the bias point of maximum drive current. Figure 9, which shows the bulk-to-gate current density along the channel for $V_{GS}=V_{DS}=V_{DD}=1$ V clearly illustrates this fact. The small increases of the gate current at the channel beginning and channel end are due to carrier diffusion from the source and drain wells.

However, this conclusion is valid provided that the physical data acquired during MC device simulation and used in gate current calculations are precise enough to give good estimates of the gate leakage. The oxide electric field is obviously not questionable, but it is not the same for the electron flow at the Si/SiO$_2$ interface. Indeed, a statistically meaningful result cannot be achieved in a MC simulation unless a very large number of sample electrons is simulated. A gate current of 1.6 A/cm$^2$ through a gate oxide surface of (0.1 $\mu$m)$^2$, for example, only corresponds to one electron per nanosecond, whereas the drive current corresponds to about 10$^5$ electrons in the same time. It means that such a gate current may be due to only one electron out of 10$^5$ per ns, which had “enough” perpendicular energy, having suffered only some scattering mechanisms, to be easily injected into SiO$_2$: this is the so-called “lucky electron.” That is the reason why a MC simulation of less than some tens of picoseconds has a high probability of not describing such scarce events. We are thus not sure that the decrease of gate current observed when $V_{DS}$ is raised for a constant and positive gate bias is not due to a poor description of high energy tails. Fortunately, the procedure described in Sec. II gives a very simple means to clarify this doubt: in order to estimate the contribution of hot carriers on the bulk to gate current of the previously described $n$-MOSFET, we arbitrarily add 20% of hot electrons ($E_t=1$ eV) to the last interface area (near drain) electron flow acquired during the MC simulation for $V_{GS}=V_{DS}=V_{DD}=1$ V. Such a perpendicular energy may be that of the ballistic electrons which suffer an elastic mechanism just at the channel end, scattering their wave vector perpendicularly to the Si/SiO$_2$ interface. This scenario well corresponds to rare events. We then obtain a gate current of 1.58 A/cm$^2$ (1.1 nA/ $\mu$m) instead of the 1.44 A/cm$^2$ (1 nA/ $\mu$m) previously obtained for the same bias point (see Fig. 7). This rather small change can be attributed to the electron concentration decrease along the channel and to the barrier potential shape effect. This confirms that the carrier interface energy distributions acquired thanks to MC device simulation contain a meaningful statistical information, and this validates the gate current calculations and results presented above.

Finally, it can be concluded from Sec. III that the maximum of gate current is unfortunately obtained for the $(V_{GS}=V_{DD}, V_{DS}=0)$ static point of complementary metal oxide semiconductor (CMOS) inverters. The gate current order of magnitude obtained for this bias point (at least some A/cm$^2$) raises a severe problem of standby power consumption for CMOS circuits based on ultrathin gate oxide MOSFET transistors.

IV. RESULTS WITH A NONUNIFORM GATE OXIDE

In all previous gate current calculations, a constant gate oxide thickness has been assumed. In spite of a very good process control, this assumption may be too optimistic. Because of the strong dependence of the transmission probability on the oxide thickness, the SiO$_2$ film nonuniformity may have a big influence on the total gate current, and probably a drastic impact on the local current density. This problem has been experimentally tackled in a recent paper, which has confirmed good reliability characteristics of 1.5 nm direct tunneling gate oxides. The aim of this section is to clarify the influence of the oxide-film nonuniformity on gate currents.

The calculation procedure described in Sec. II can be simply improved by adding a description of the transistor width $W$. $W$ is assumed to be divided into $n_W$ zones. The gate oxide surface is now meshed into $n_L \times n_W$ surface areas (see Fig. 2). The oxide thickness of each area is randomly chosen according to a Gaussian law of standard deviation $\sigma$, with a maximum deviation authorized $\Delta t_{ox}$ to the mean value $t_{ox}$. All other physical data (oxide field, carrier flow) are assumed to be the same for all areas having the same channel length index $l_x$. By summing over all areas, gate currents and surface distributions are then calculated with Eqs. (1) and (2). Actually, no new physical information is added, except the oxide-film nonuniformity.

We now apply this to the direct tunneling MOSFET of Sec. III, which is assumed to have a width of 200 nm (see Fig. 2). The mean oxide thickness is $t_{ox} = 1.5$ nm, the standard deviation and the maximum deviation assumed are, respectively, $\sigma = 0.18$ nm and $\Delta t_{ox} = 0.4$ nm (on the basis of Ref. 3). Gate currents tend to limit values when the number $n_L$ of areas in the transistor width is increased up to 100. The gate current over drain voltage characteristic for $V_{GS}=V_{DD}=1$ V when oxide layer thickness fluctuations are taken into account is shown in Fig. 7 (dashed line): a gate current of 32.3 A/cm$^2$ (22.6 nA/ $\mu$m) is obtained for $V_{GS}=V_{DD}=1$ V and $V_{DS}=0$, and a gate current of 7.8 A/cm$^2$ (5.45 nA/ $\mu$m) for $V_{GS}=V_{DS}=V_{DD}=1$ V. We see that a correct agreement is now achieved with experimental data when technological fluctuations of the oxide thickness are included: measured gate current values from 10 to 45 A/cm$^2$.
and for devices comparable to the one studied in this article are reported in Refs. 1–3. No fit of the oxide mass, which is then thickness dependent, is needed to. From our point of view, the oxide film nonuniformity description is more important than the controversy image potential correction. Moreover, the model shows that the maximum of gate current is no more linked to the carrier energy peak in the channel but is more linked to the oxide weak points. Moreover, Figs. 10(b) and 10(c), respectively, show that charge injection occurs near source for $V_{GS}=V_{DS}=V_{DD}=1 \text{ V}$, and near drain for $V_{DS}=0$ and $V_{DD}=V_{DS}=1 \text{ V}$.

The gate current calculation method also allows a simple estimate of the oxide surface through which flows the major part of the gate current. To do this, the following treatment is applied: gate surface areas are arranged in decreasing gate current densities. Then, the gate current which flows through the first “i” areas and the corresponding gate surface are estimated. By normalizing these quantities with, respectively, the total gate current and the total gate surface and repeating it from $i=1$ to $n_{G}\times n_{W}$, we obtain an information about the gate current surface distribution. Figure 11 presents the results obtained for $V_{GS}=V_{DD}=1 \text{ V}$ and $V_{DS}=0$ with a constant oxide thickness (solid line) and with a nonuniform oxide layer (dashed line). We see in the latter case that about 80% of the current flows through about 10% of the surface. This fact may explain nonreproducible current versus voltage characteristics of very small area capacitors.

**V. CONCLUSION**

In this article a gate current calculation procedure using Monte Carlo device simulation has been applied to a 0.07 $\mu$m gate length $n$-MOSFET with a direct tunneling 1.5 nm gate oxide. In a first step gate currents have been calculated assuming a uniform gate oxide thickness layer for different gate and drain voltages. A sharp dissymmetry between gate currents for negative and positive gate voltages and for a zero drain voltage has been explained by a tunneling potential barrier effect more favorable to injection from bulk for positive gate bias. It has been shown that the drain voltage influence on gate currents of ultrathin gate oxide $n$-MOSFETs for $V_{GS}=V_{DD}$ differs from the one for thick gate oxide transistors: the maximum of gate current is no more linked to the carrier energy peak in the channel but is near the source where the electron concentration is the largest. Nevertheless, the main conclusion is that the maximum of gate current corresponds to one of the two static points of CMOS inverters ($V_{GS}=V_{DD}$, $V_{DS}=0$), which raises a severe problem of standby power consumption. A simple modeling of the oxide film nonuniformity has then given a correct agreement with published experimental gate currents, without fitting the oxide mass or including the classical image potential correction. Moreover, the model shows that the oxide layer locally undergoes very high current densities in

![Figure 10](image1.png)

**FIG. 10.** Surface distribution of gate current density: 10 (a): for $V_{GS}=1 \text{ V}$ and $V_{DS}=0$: no preferential injection area, 10 (b): for $V_{GS}=V_{DS}=1 \text{ V}$: injection near source, 10 (c): for $V_{GS}=0$ and $V_{DS}=1 \text{ V}$: injection near drain.

![Figure 11](image2.png)

**FIG. 11.** Surface distribution of gate current density for $V_{GS}=V_{DS}=V_{DD}=1 \text{ V}$, and near drain for $V_{DS}=0$ and $V_{DD}=V_{DS}=1 \text{ V}$.
the vicinity of weak points where the oxide thickness is small. Even if microscopic degradation phenomena are up to now not very well known, it is obvious that such weak points are linked to the soft breakdown mechanism of the ultrathin oxide layer.