Nanoscale field-effect transistors: An ultimate size analysis

F. G. Pikus and K. K. Likharev
Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11784-3800

(Received 16 May 1997; accepted for publication 21 October 1997)

We have used a simple, analytically solvable model to analyze the characteristics of dual-gate metal-oxide-semiconductor field-effect transistors (MOSFETs) with 10 nm-scale channel length L. The model assumes ballistic dynamics of two-dimensional electrons in an undoped channel between highly doped source and drain. When applied to silicon n-MOSFETs, calculations show that the voltage gain (necessary for logic applications) drops sharply at $L\sim 10$ nm, while the conductance modulation remains sufficient for memory applications until $L\sim 4$ nm. © *1997 American Institute of Physics.* [S0003-6951(97)02151-7]

The recent industrial forecast[1] predicts that the progress in scaling down the metal-oxide-semiconductor field-effect transistors (MOSFETs) will result in MOSFETs with channel length L at least as short as 70 nm. Such devices would allow the implementation of dynamic random-access memories (DRAMs) with density up to ~ 5 Gb/cm^2.

On the other hand, single-electron transistors (SET), in particular the recently suggested SET/FET hybrids,[2] may allow room-temperature operation of dynamic memories with a density of 100 Gb/cm^2 and beyond, using a silicon-based technology with minimum feature size below ~ 5 nm. Thus, it is very important to understand whether purely MOSFET-based devices can operate on a comparable scale. Recent experiments[3] with ~ 10 nm-long MOSFETs have shown that their conductance can be gate modulated by at least three orders of magnitude. We are not aware, however, of any published analysis of 10 nm-scale MOSFETs, with the exception of a couple of Monte-Carlo-calculated $I-V$ curves for some particular devices with $L=15$ nm (Ref. 4) and $L=30$ nm (Ref. 5). The main goal of this work was to use a simple model of nanoscale MOSFETs, which captures the essential physics of such devices, for semianalytical calculation of their basic characteristics. All the examples presented below are for silicon-based n-MOSFETs, although our approach is certainly more general.

For nanoscale devices channel doping becomes *unacceptable*. In fact, the volume of the channel region of a 10 nm-scale transistor is of the order of 2×10^{-19} cm^3, so that even doping at a level as high as 10^{20} cm^3 (beyond the degeneracy threshold for silicon) would result in less than 20 dopants in the whole channel, and hence in large statistical device-to-device variations of transistor parameters. For the same reason, contact doping has to be high ($>3\times 10^{20}$ cm^{-3}). On the other hand, channel doping is *unnecessary*, because the channel length is comparable to the screening length in highly doped source and drain, so that the carriers may be delivered from the contacts.[6] This is why in our model the channel is a layer of an intrinsic semiconductor connecting n^+ source and drain [Fig. 1(a)]. Because of the absence of impurities, electron scattering is so small at our scale of channel length ($L\sim 10$ nm) that it may be ignored,[3,4] and the electron transport in the channel considered as completely ballistic. Electrons in the source and drain are assumed to be in thermal equilibrium.

Thickness $2s$ of the channel is assumed to be so small that in fact it presents a quantum well with two-dimensional electron gas. For Si (100) surface the sixfold valley degeneracy of the bulk Si is lifted and only two valleys participate in the lateral transport,[7] unless the carrier density is so large that the electrons start to populate higher subbands (which is not the case for the devices considered below). The channel is sandwiched between plates of a "dual gate,"[8,9] the optimal structure to suppress the "short channel effects" (for a review see, e.g., Ref. 10). If $2s$ is small enough, z dependence of the electric potential ϕ can be assumed quadratic inside the channel and linear inside the dielectric (the well-known parabolic approximation[11]). Substituting this dependence into the Poisson equation one arrives at a one-dimensional equation for the distribution of electric potential ϕ at $z=0$ along the length of the device:

$$\frac{d^2\phi}{dx^2}-\frac{\phi-V_g}{\lambda^2}=-\frac{4\pi\rho(x)}{\kappa_1}. \tag{1}$$

Here $\rho=-en$ is the electric charge density averaged over the channel thickness $2s$; within source and drain ρ includes also the dopant charge $+eN_D$. V_g is the gate-source voltage, and λ is the effective screening length

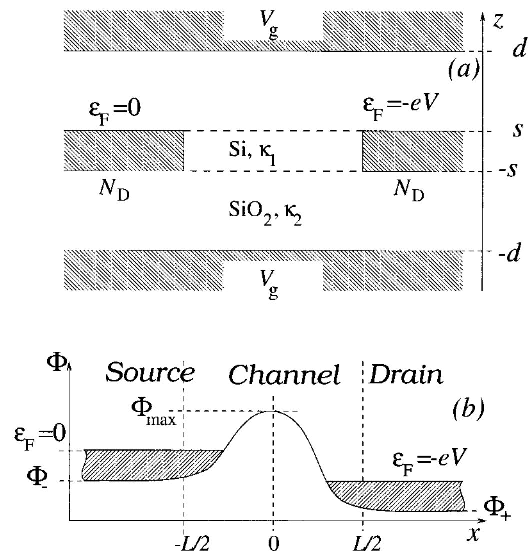

FIG. 1. (a) Sketch of the MOSFET model considered in this work and (b) scheme of the conduction band edge diagram for the electron potential energy $\Phi=-e\phi$.

$$\lambda^2 = \frac{s^2}{2} + \frac{\kappa_1}{\kappa_2} s(d-s). \tag{2}$$

For the examples presented below we took the dielectric constants to be $\kappa_1 = 12$ (Si) and $\kappa_2 = 4$ (SiO_2). The parabolic approximation is strictly valid[11] if λ is much larger than s but much less than relevant x-scales of the problem, notably, the channel length and the screening length λ_2 of the electron gas without the ground plane. λ_2 is close to the effective Bohr radius $a_B = \alpha\kappa\hbar^2/me^2$, where $\alpha \approx 1.5$ accounts for valley degeneracy.

The second equation relating ρ and ϕ follows from the condition of conservation of the ballistic current components $j_\epsilon^\pm$ for each energy ϵ:

$$j_\epsilon^\pm(x) = -en_\epsilon^\pm(x)v_\epsilon^\pm(x) = \text{const} = j_{\epsilon 0}^\pm, \tag{3}$$

where $v^\pm(x) = \pm\sqrt{2[\epsilon - \Phi(x)]/m}$ is the x component of the electron velocity, $\Phi(x) = -e\phi(x)$ is the electron potential energy, and $j_{\epsilon 0}^\pm$ are currents into the ballistic channel from the surfaces of source (sign $+$) and drain (sign $-$). The latter currents can be found from the usual thermal equilibrium distribution, then the velocity $v_\epsilon(x)$ and density $n_\epsilon(x)$ are found with self-consistent values of Φ. The total electron density n and current j (per unit channel width) can be obtained as integrals of $[n_\epsilon^+(x) + n_\epsilon^-(x)]$ and $(j_{\epsilon 0}^+ - j_{\epsilon 0}^-)$, respectively, over energies $\epsilon > \Phi(x)$ [or $\epsilon > \Phi_{max}$ if the potential maximum lies between the point x and the contact, Fig. 1(b)]. For the parameters considered below, tunneling under the potential barrier[12] can be ignored for $L \geqslant 5$ nm.

The resulting simple set of equations allows the channel length L and current j to be expressed explicitly as analytical (though bulky) integrals for an arbitrary relation between temperature T and Fermi energy ϵ_F in the source and drain, provided that Φ_{max} is considered known (together with the real parameters of the system, including the gate and source-drain voltages). The resulting function $L = L(\Phi_{max})$ can be numerically interpolated to the desired values of channel length.

Figures 2–4 show the results of such semianalytical calculations for a Si/SiO_2 n-MOSFET with a contact doping level of $N_D = 3\times10^{20}$ cm^{-3}, channel thickness $2s = 1.5$ nm, and gate oxide thickness $d - s = 2.5$ nm. For these parameters, the parabolic approximation is indeed applicable: Eq. (1) yields $\lambda \approx 2.5$ nm. Though the factor $\lambda/a_B \approx 1.5$ can be hardly considered as large, we still believe that the parabolic approximation should work reasonably well provided that the channel length is much larger than a_B, because the screening at a_B is polynomial[5] (for a point charge, $\phi \sim r^{-3}$ at $r \gg a_B$), while that due to the gates is exponential. Hence in a relatively long device ($L \gg a_B$) the wave vectors of the order of a_B^{-1} can give only a small correction to our results, crudely equivalent to a channel length uncertainty of the order of $a_B \sim 2$ nm.

Figure 2 shows the distribution of electric potential (solid lines) and electron density (dashed lines) along a 10 nm-channel MOSFET. Clearly, our model is capable of describing the penetration of the electric field into source and drain, as well as the pinch-off effect in relatively long devices (such as that shown in Fig. 2). The $I-V$ curves of such devices show clear saturation (Fig. 3) and hence relatively high voltage gain $G_V = dV/dV_g|_{I=\text{const}}$. Their transconductance is also as high as that of the best present-day transistors,[10] for example $S \approx 1.5$ S/mm at $V = V_g = 0.5$ V for $L = 10$ nm.

However, as the length L becomes comparable to the effective screening length λ, electron concentration in the

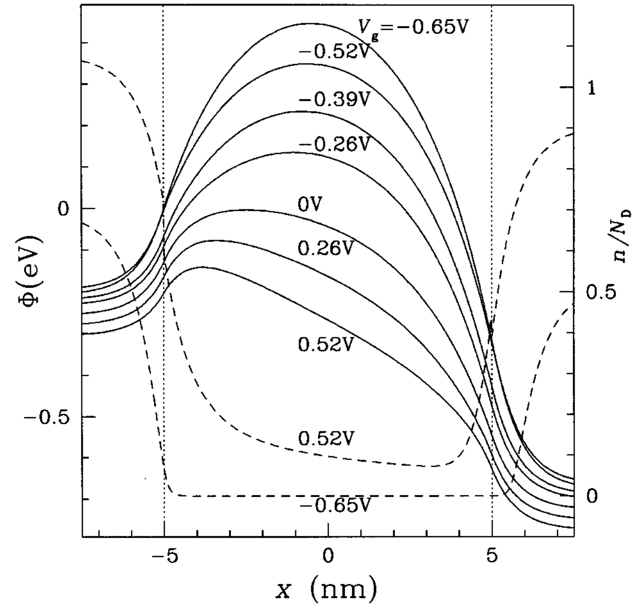

FIG. 2. Distribution of electron potential energy Φ (solid lines) and density n (dashed lines) along a two-dimensional silicon n-MOSFET with a 10 nm-long intrinsic channel and n^+ source and drain ($N_D = 3\times10^{20}$ cm^{-3}), for a moderate negative bias $V = -0.52$ V and several values of gate voltage V_g. Geometric parameters [Fig. 1(a)] are: $2s = 1.5$ nm and $2d = 6.5$ nm. Temperature $T = 300$ K.

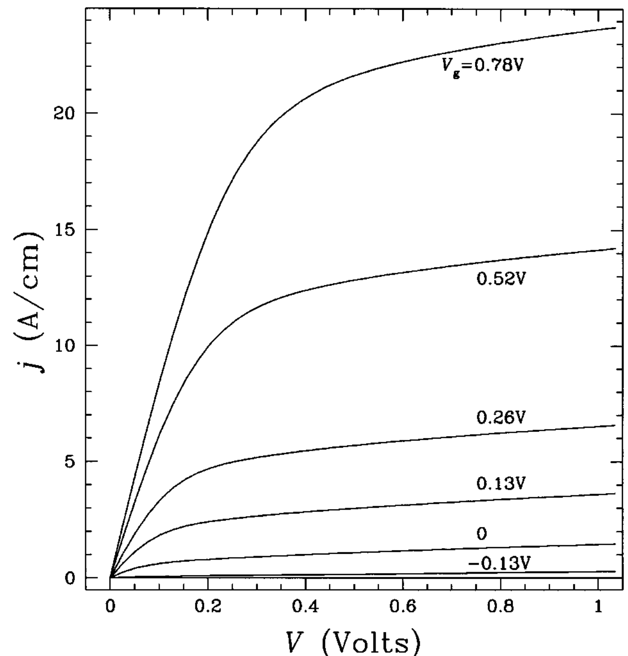

FIG. 3. Source-drain $I-V$ curves for a n-MOSFET with $L = 10$ nm for several values of gate voltage V_g. Device parameters are the same as in Fig. 2.

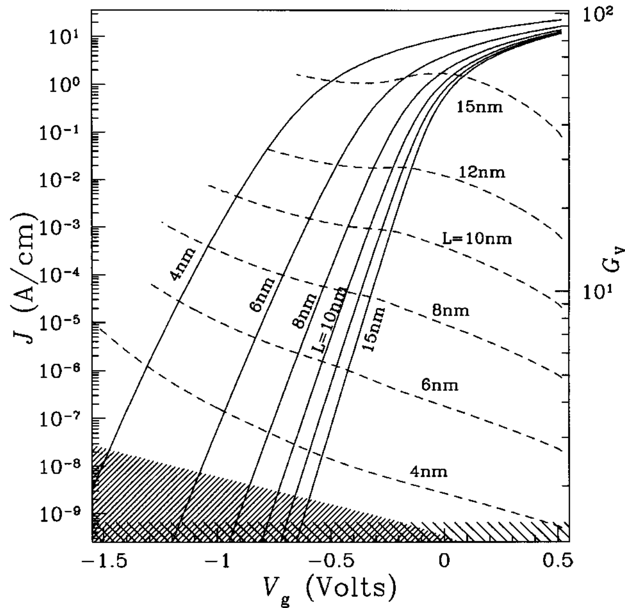

FIG. 4. Linear current density j (solid lines) and voltage gain $G_V = dV/dV_g|_{I=\mathrm{const}}$ (dashed lines) as functions of gate voltage V_g for various channel lengths L and source-drain voltage near the onset of saturation ($V = 0.52$ V). The fine hatching shows the area of parameters where the gate leakage current exceeds the drain current. The coarse hatching shows the region where the intrinsic carriers in the channel cannot be ignored.

channel becomes controlled more by the drain voltage and less by the gate. The saturation disappears, and the voltage gain above the threshold drops sharply, especially in the range of V_g where the current is substantial (Fig. 4). Beyond ≈ 10 nm logic applications of nano-MOSFETs become problematic. However, for DRAM applications (see, e.g., Ref. 13) the voltage gain, as well as the depletion operation mode of the transistors ($V_t < 0$), are of minor importance, since the compact memory cells can be controlled by drivers using more conservative technology. What is really important for DRAMs is to have the channel current modulated by at least eight to nine orders of magnitude (this determines the necessary ratio of retention refresh time to read/write time).

The modulation is limited from the side of small currents by two major effects not accounted for in our model: the thermal activation of holes and tunneling through the gate oxide. The hole activation (and hence the sharp loss of gate control) takes place when the maximum potential in the channel approaches the middle of the band gap; the corresponding region in Fig. 4 is coarsely hatched. To estimate the tunneling effects, we have calculated the current taking into account the image charge effects and using potential barrier height of 3.2 eV and effective electron mass of SiO_2 $0.4m_0$ (see, e.g., Ref. 14), and assuming that the gates overlap source and drain by 2 nm (the final conclusions are fairly insensitive to this value). Fine hatching in Fig. 4 shows the region where the tunneling current becomes larger than the channel current. The boundary of this region rises very rapidly with a decrease in oxide thickness; for $d-s=2$ nm it corresponds to $j \sim 10^{-6}$ A/cm. On the other hand, if the oxide is thicker than 2.5 nm, the transconductance and voltage gain fall, while the modulation range remains virtually the same because of the hole activation effects. Hence this value of oxide thickness may be considered as optimal. Thus the channel current modulation can hardly be better than that shown in Fig. 4. One can see that the maximum modulation depth falls below 3×10^8 at $L \approx 4$ nm; crudely, this value may be considered as the minimum length of silicon-based MOSFETs for application in traditional DRAM cells. This does not preclude the possibility that smaller transistors could be used in some novel memories based on different physical principles.

In our analysis, several other potentially important effects have been neglected, including impact ionization, finite rate of the energy relaxation, and source and drain resistance and self-heating. Simple estimates show, however, that all these effects can be ignored for our parameters.

To summarize, we have found that for silicon-based MOSFET room-temperature operation with reasonable parameters is still feasible for channel length L down to ~ 10 nm for logic circuits and ~ 4 nm for DRAM cells. Since we have considered a virtually optimal MOSFET structure, it is hard to imagine that these limits could be surpassed without suggesting some radically new physical ideas.

Fruitful discussions with D. Ferry, M. Fukuma, S. Luryi, V. Mitin, M. Shur, and J. Tucker are gratefully appreciated. The work was supported in part by ONR/DARPA within the framework of the Ultra Electronics program.

[1] *The National Technology Roadmap for Semiconductors* (Semiconductors Industry Association, San Jose, CA, 1994).
[2] K. K. Likharev and A. N. Korotkov, in *Proceedings of the 1995 International Semiconductor Device Research Symposium*, (University of Virginia, Charlottesville, VA, 1995), p. 335.
[3] L. Guo, E. Leobandung, and S. Chou, Appl. Phys. Lett. **70**, 850 (1997).
[4] M. Fukuma, in *Digest of Technical Papers*, Symposium on VLSI Technology, San Diego, CA, 10–13 May 1988 (unpublished), pp. 7 and 8.
[5] D. J. Frank, S. E. Laux, and M. V. Fischetti, IEEE Trans. Electron Devices **40**, 2103 (1993).
[6] A. A. Grinberg and S. Luryi, J. Appl. Phys. **61**, 1181 (1987).
[7] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[8] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, IEEE Electron Device Lett. **8**, 410 (1987).
[9] T. Tanaka, K. Suzuki, H. Horie, and T. Sugii, IEEE Electron Device Lett. **15**, 386 (1994).
[10] Y. Taur, D. A. Buchanan, W. Chen, D. J. Frank, K. E. Ismail, S.-H. Lo, G. A. Sai-Halasz, S. G. Viswanathan, H.-J. C. Wann, S. J. Wind, and H.-S. Wong, Proc. IEEE **85**, 486 (1997).
[11] K. K. Young, IEEE Electron. Dev. **36**, 504 (1989); V. Aggarwal *et al.*, Solid-State Electron. **37**, 1537 (1994).
[12] J. Tucker, C. Wang, and P. S. Carney, Appl. Phys. Lett. **65**, 618 (1994).
[13] B. Prince, *Semiconductor Memories*, 2nd ed. (Wiley, Chichester, 1991).
[14] B. Brar, G. D. Wilk, and A. C. Seabaugh, Appl. Phys. Lett. **69**, 2728 (1996).