Long-Range Coulomb Interactions in Small Silicon Devices: Transconductance and Mobility Degradation

M. V. Fischetti* and S. E. Laux**

IBM Research Division, Thomas J. Watson Research Center
P. O. Box 218, Yorktown Heights, NY 10598, USA
*fischet@watson.ibm.com **laux@watson.ibm.com

ABSTRACT

In small silicon devices, conduction electrons in the channel are subject to long-range Coulomb interactions with electrons in the heavily-doped drain, source, and gate regions. We show that for devices with channel lengths shorter than about 40 nm and oxides thinner than 2.5 nm these interactions cause a reduction of the electron velocity. We present results obtained using both semiclassical two-dimensional self-consistent Monte Carlo-Poisson simulations and a quantum-mechanical model based on electron scattering from gate-oxide interface plasmons.

Keywords: Monte Carlo simulations, Coulomb interactions, MOS transistors, mobility, VLSI scaling

1 INTRODUCTION

Electrons in the conducting channel of small metal-oxide-semiconductor field-effect-transistors (MOSFETs) are in proximity to heavily-doped regions with free electron densities exceeding 10^{20} cm^{-3}: The source (S) and drain (D) regions, separated by a few tens of nanometers, and the gate (G) region, separated from the channel by as little as 1.5 nm of SiO_2[1]. In this letter we show that long-range Coulomb interactions between the channel and the heavily-doped regions, while of marginal importance in present-day devices, will have a negative impact on the performance of devices with channels shorter than 40 nm and oxides thinner than 2.5 nm.

We start by reviewing some properties of a high-density electron gas and discuss how the two-dimensional (2D) self-consistent Monte Carlo/Poisson simulations we have employed[2] can reproduce the three-dimensional (3D) semiclassical behavior.

2 HIGH-DENSITY ELECTRON GAS

Within the semiclassical 'modified Hartree model'[3] the electron gas is described as a uniform jellium of negative charge embedded in a background of point-like positive charges. Coulomb interactions lower the total energy of the system by an amount

$$\Delta E \approx -1.451\frac{e^2 n^{1/3}}{4\pi\epsilon_s}\ , \qquad (1)$$

where e is the magnitude of the electron charge, n the electron density, and ϵ_s the static, long-wavelength dielectric function of the semiconductor. By the virial theorem, this change of total energy results from a lowering of the time-averaged potential energy $\Delta U = 2\Delta E$ and an increase of the time-averaged kinetic energy, $\Delta K = -\Delta E$, corresponding to the additional 'agitation' of the electrons as they repel each other. The (semi)classical picture obviously misses exchange effects (Hartree-Fock) – which in Si constitute an additional 26% reduction of the total energy – and correlation effects, which are negligible in the density range of interest[4]. The Coulomb-induced kinetic energy causes the electron distribution in **k**-space to deviate from the equilibrium Fermi-Dirac distribution: A Fermi-Dirac distribution in *total* energy must be convoluted with a fluctuating potential, resulting in an electron distribution in *kinetic* energy which exhibits stronger high-energy tails. The fluctuating potential can be viewed as due to zero-point and thermal fluctuations associated with the collective excitations of the electron gas, *i.e.*, plasmons. Quantum mechanically, the canonical quantization procedure[5] yields for the field associated with a plasmons of wave vector $\mathbf{q}$ and frequency $\omega_p = [e^2 n/(\epsilon_s m_c)]^{1/2}$ (where m_c is the electron conductivity mass) an amplitude $\phi_{\mathbf{q}} = -(i/q)[\hbar\omega_p/(2\epsilon_s)]^{1/2}$, so that the r.m.s. fluctuations associated with all plasmons up to the cutoff wave vector q_c[6] will be:

$$< \phi >_{QM} = (1+2n_p)^{1/2} \left(\frac{\hbar\omega_p q_c}{4\pi^2\epsilon_s}\right)^{1/2}\ , \qquad (2)$$

where n_p is the thermal plasmon number. Semiclassically, the potential fluctuations can be estimated by summing the time-averaged electrostatic energy density $(1/2)\phi_{\mathbf{q}}\rho_{\mathbf{q}}$ (where $\rho_{\mathbf{q}}$ is the polarization charge associated with a mode $\mathbf{q}$) over all modes $\mathbf{q}$, and equating the result to the time-averaged kinetic-energy density $n\Delta K$ given above, obtaining:

$$< \phi >_{SC} = \left(\frac{4.353}{\pi}\right)^{1/2} \frac{e^2 n^{2/3}}{\epsilon_s \beta}\ , \qquad (3)$$

where β is the screening wave vector, of the order of the Landau-damping parameter. When using 2D self-consistent Monte Carlo/Poisson simulations, a proper

selection of the mesh spacing, Δx, and statistical weight s of the superparticles (in charge/unit-length) guarantees the equivalence of the 2D simulations with the 3D semiclassical results: By chosing $\Delta x \sim 1/\beta$ and $s \sim \beta$ we recover both the desired semiclassical value for ΔK as well as the 'correct' amplitude of the potential fluctuations induced by collective excitations. In the density-range of interest ($n \approx 10^{19} - 10^{20}$ cm^{-3}), $< \phi >_{SC}$ is larger than $< \phi >_{QM}$ by no more than a factor of $\sqrt{2}$. Considering the 'fuzziness' intrinsic to the definition of the parameter q_c (here taken to be the classical value β), we see that quantum corrections are minor and that 2D simulations perform accordingly to what is expected from 3D semiclassical physics. Additionally, we have verified that the 2D model is able to yield the correct wavelength and frequency dependence of the dielectric response, the correct frequency dependence of the potential fluctuations (peaked around the plasma frequency), and the approximate value of band-gap narrowing effects.

3 MOS SCALING STUDY

The main point of our work can now be simply stated: The potential fluctuations present in the S/D and G regions extend into the channel. For channel lengths comparable to the screening length of the channel, the S/D plasma fluctuations cause a rapid 'randomization' of the electron energy distribution, similarly to the effect caused by short-range interactions near the drain[7], resulting in more carriers at higher energy, and so in additional momentum loss via phonon scattering, and in a lower electron velocity. Quantum-mechanically, this process may be regarded as initiated by emissions/reabsorptions of drain-plasmons by channel electrons. On the other hand, the excitation and absorption of plasma modes in the gate results in a net momentum loss for the conduction electrons, and thus in a reduced drain current.

To quantify these statements, we have employed our 2D, self-consistent Monte Carlo/Poisson program to simulate n-channel MOSFETs with metallurgical channel lengths ranging from 100 nm to 11.7 nm. The oxide thickness varies linearly with channel length from 5.6 nm (100 nm channel) to 0.7 nm (11.8 nm channel). Physical dimensions, junction depths, and channel doping are scaled as required by conventional scaling[8]. S/D and G doping levels, already at the technological limit in the 'template' 50 nm-long device[9], are kept constant. S/D and S/G biases are chosen around 1 V in all cases. The physical models employed have been described in the past: Refs. [2] and [10] (Appendix A) describe the electron-phonon and short-range Coulomb scattering models; Ref. [12] describes transport in inversion layers. Two major deviations from the model of Ref. [12] should be noted: 1. Results from Ref. [11] have prompted the use of different anisotropic intervalley deformation potentials for scattering with acoustic phonons and also intervalley deformation potentials, now from Ref. [14]. The discouraging results of Ref. [12] regarding the phonon-limited mobility are now much improved[13]. 2. Short-range electron-electron scattering in inversion layers is now included improving on the model by Lee and Galbraith[15] by accounting for the anti-symmetrized matrix element, dynamic multi-subband screening, and using the Green function suitable to the MOS system under study (obtained following the appendix of Ref. [16]). Finally, we have accounted for short-range scattering between channel and gate electrons, but have neglected scattering with remote ionized impurities in the gate depletion layer[17].

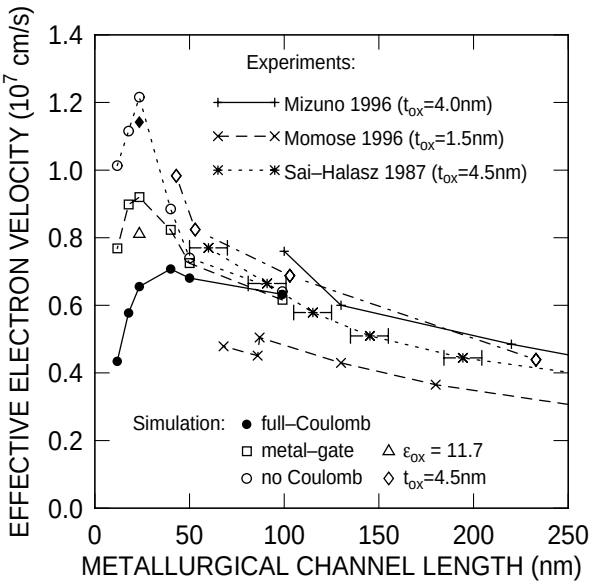

Figure 1: Room temperature effective electron velocity obtained from two-dimensional Monte Carlo simulations of n-MOSFETs scaled from a 'nominal' 50 nm channel-length/2.8 nm-thick oxide to smaller (11.8/0.7 nm) and larger (100/5.6 nm) devices. Results obtained accounting for Coulomb interactions with the S/D and G regions (dots, 'full Coulomb') are compared to results obtained by ignoring the channel-gate interaction (open squares, 'metal gate'), and all long- and short-range Coulomb interaction (circles, 'no Coulomb'). Results for a 23.5/1.4 nm device with a high-dielectric-constant insulator (triangle), and simulation results for 4.5 nm-oxide devices are also shown. The calculated results are no more than 10% accurate, because of numerical noise and since g_m is evaluated from the difference in calculated drain currents at G/S biases of 1.0 V and 0.75 V. Finally, comparison is made with some published experimental data.

Figure 1 shows the simulated 'effective' electron velocity (defined as transconductance, g_m, divided by gate-channel capacitance, C_{gc}) at 300 K as a function of device dimensions. A maximum 'speed' is obtained for a channel length between about 40 nm ($g_m \approx 1000$ S/m, $t_{ox} = 2.1$ nm) and 23.5 nm ($g_m \approx 1700$ S/m, $t_{ox} =$

1.4 nm). At the minimum simulated dimensions (11.8 nm channel length, t_{ox} = 0.7 nm), the device is probably 'unrealistic' and semiclassical transport is likely to be inapplicable. Yet, within the theoretical and practical assumptions, its performance lies well below any 'ballistic limit'[18]. Coulomb interactions depress the electron velocity by more than a factor of 2, S/D-channel and G-channel interactions sharing the blame almost equally[20]. More conventional scattering mechanisms (surface-roughness and scattering with acceptors in the channel) are responsible for the remaining performance degradation seen in the 'no Coulomb' results at lengths below 23.5 nm. We also show our older simulation results[19] (no channel/G interactions) for devices having a 4.5 nm-thick oxide, as well as some experimental data from Refs. [21] and [1] (having extracted the effective velocity from the published g_m) as well as from Ref. [22]. Finally, note how replacing SiO_2 with a dielectric having the permittivity of Si mitigates the negative effect of gate plasmons.

4 INTERFACE PLASMONS AND EFFECTIVE MOBILITY

In order to support the semiclassical, 2D model with more accurate quantum-mechanical calculations, we have considered the interaction between channel electrons and 2D surface plasmons (SP) localized at the gate-oxide interface. We have calculated the dispersion of the interface modes for a bulk-Si/SiO_2/inverted-Si system following Ngai and Economou[5] using the long-wavelength dielectric functions $\epsilon_s(1-\omega_g^2/\omega^2)$ for the bulk-Si gate, $\epsilon_s(1-\omega_{2D}(\mathbf{q})^2/\omega^2)$ for the inverted Si channel, where $\omega_{2D}(\mathbf{q})^2 \approx \sum_i e^2 n_i q/(\epsilon_s m_{c,i})$, the sum extending over all subbands (and ladders) i with density n_i and conductivity mass $m_{c,i}$, as it follows from the longitudinal expression of Dahl and Sham[16]. The oxide TO-phonon modes have been neglected and we have employed a constant ϵ_{ox}. In the non-retarded limit, the field, $\phi_q^{(r)}(z)$, associated to the r-th mode of wave vector q takes the form:

$$\phi_q^{(r)}(z) \;=\; \frac{\epsilon_{ox}-\epsilon_s}{\epsilon_{ox}+\epsilon_s}\left(\frac{\hbar\omega_q^{(r)}}{2qD_q}\right)^{1/2}\exp(-qz)\;, \qquad (4)$$

in the channel ($z>0$) where

$$D_q \;=\epsilon_s+\epsilon_{ox}[1-e^{-2qt_{ox}}]\left(\frac{\epsilon_{ox}-\epsilon_s}{2\epsilon_{ox}}\right)^2+$$
$$\epsilon_{ox}[e^{2qt_{ox}}-1]\left(\frac{\epsilon_{ox}+\epsilon_s}{2\epsilon_{ox}}\right)^2+\epsilon_s e^{-2qt_{ox}}\left(\frac{\epsilon_{ox}-\epsilon_s}{\epsilon_{ox}+\epsilon_s}\right)^2 (5)$$

t_{ox} being the oxide thickness. Since modes localized at the channel-insulator interface do not subtract momentum from the electron gas, we have considered only the fraction of the SP field associated with modes localized at the gate-insulator interface. subband dielectric function of the Si inversion layer given in Ref. [16], and modeling the channel as a triangular well with a depletion charge of 5.5×10^{11} cm^{-2}. Other momentum relaxation channels are as described previously for the Monte Carlo simulations. Finally, we have calculated the effective electron mobility as a function of electron density in the channel using the Kubo-Greenwood expression. Figure 2 shows the effect of gate-oxide SPs: Given the uncertainty concerning the exact definition of q_c, we present results obtained using two alternative cutoffs: When accounting only for plasmons of wave vector smaller than the zero-temperature Landau-damping parameter in the gate, we effectively suppress the low-energy acoustic branch. High electron densities are required before the higher-energy optical modes can be felt. When considering all modes undamped up to the classical value β (presumably a more appropriate cutoff at 300 K), the mobility is depressed also at lower electron sheet densities. We have also found that the mobility increases very quickly to its (phonon and surface roughness-limited) thick-oxide value for t_{ox} exceeding about 3 nm, and it also increases with increasing ϵ_{ox}.

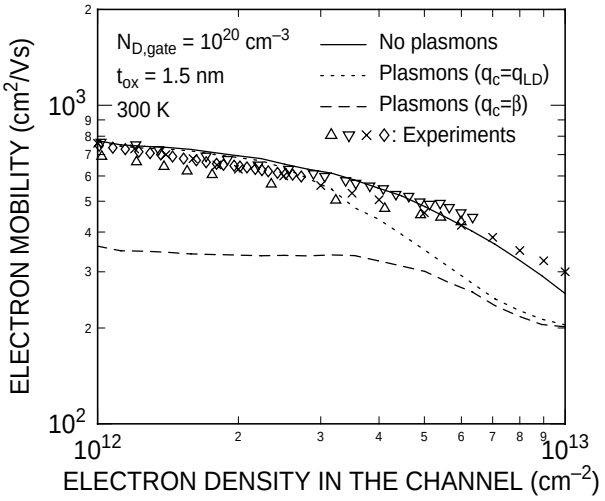

Figure 2: Room temperature effective electron mobility in the channel of a n-MOSFET calculated accounting only for phonon and surface-roughness scattering (solid line), and including scattering with gate-oxide interface plasmons across an oxide 1.5 nm-thick damped at the zero-temperature Landau-damping wave vector (dotted line) or at the Thomas-Fermi screening parameter (dashed line). A selection of experimental data for thick oxides (same symbols as in Ref. [12]) is also shown.

5 CONCLUSIONS

The results of the preceeding section show the same qualitative trends seen in our semiclassical simulations,

which apply mainly to higher-energy transport and self-consistently account for plasmon damping effects and for other modes present in the device. On the other hand, the simpler quantum-mechanical mobility calculations apply only to Ohmic transport, but account for the 'correct' thermal occupation of the plasmons and for the 'correct' quantized energy-exchange between electrons and plasmons. The fact that both schemes provide qualitatively identical (and quantitatively similar) results lends credibility to our main conclusions: Further scaling of Si MOSFETs will not reward us with the increased performance we have come to expect.

REFERENCES

[1] H. S. Momose *et al.*, Int. Electron Devices Meet., Tech. Dig. IEEE, 109 (1996).

[2] M. V. Fischetti and S. E. Laux, Phys. Rev. B **38**, 9721 (1988); S. E. Laux and M. V. Fischetti, in *Monte Carlo Device Simulation: Full Band and Beyond*, Karl Hess ed. (Kluwer, Boston, Massachusetts, 1991).

[3] C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963), Chaps 5 and 6.

[4] G. D. Mahan, J. Appl. Phys. **51**, 2634 (1980).

[5] K. L. Ngai and E. N. Economou, Phys. Rev. B **4**, 2132 (1971).

[6] See, for example, D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953).

[7] M. V. Fischetti and S. E. Laux, Int. Electron Devices Meet., Tech. Dig. IEEE, 305 (1995).

[8] R. H. Dennard *et al.*, IEEE J. Solid State Circuits **SC-9**, 256 (1974).

[9] M. Hargrove *et al.*, Int. Electron Devices Meet., Tech. Dig. IEEE, 627 (1998).

[10] M. V. Fischetti, S. E. Laux, and E. Crabbé, J. Appl. Phys. **78**, 1058 (1995).

[11] M. V. Fischetti and S. E. Laux, J. Appl. Phys. **80**, 2234 (1996).

[12] M. V. Fischetti and S. E. Laux, Phys. Rev. B **48**, 2244 (1993).

[13] `http://www.research.ibm.com/DAMOCLES/html_files/mueff.html`

[14] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta, Phys. Rev. B **12**, 2265 (1975).

[15] S.-C. Lee and I. Galbraith, Phys. Rev. B **56**, 15796 (1999).

[16] D. A. Dahl and L. J. Sham, Phys. Rev. B **16**, 651 (1977).

[17] See A. Chin, W. J. Chen, T. Chang, R. H. Kao, B. C. Lin, C. Tsai, and J. C.-M. Huang, IEEE Electron Device Lett. **EDL-18**, 417 (1997) and M. S. Krishnan *et al.*, IEDM Tech. Dig., 571 (1998). Despite these claims, we think that scattering with gate-impurities, being screened by electrons in the chanel and in the charge-neutral region of the Si gate, and having a minimum impact parameter $\approx t_{ox}$, should not play a major role at high gate doping and large electron sheet densities.

[18] S. Datta, F. Assad, and M. S. Lundstrom, Superlattices and Microstructures, **23**, 771 (1998).

[19] S. E. Laux and M. V. Fischetti, IEEE Electron Device Lett. EDL-9, 467 (1987).

[20] A large momentum-transfer rate between electron layers across a dielectric has been previously obtained by C. Jacoboni and P. J. Price, Solid-State Electron. **31**, 649 (1988).

[21] T. Mizuno and R. Ohba, Int. Electron Devices Meet., Tech. Dig. IEEE, 105 (1996).

[22] G. A. Sai-Halasz *et al.*, IEEE Electron Device Lett. **EDL-9**, 463 (1987).