RC21498(96980)2JUN1999

Physical Science

Research Report

Coulomb Interactions and Hot-Electron Effects in Sub 0.1 μm Si MOSFETs

M. V. Fischetti, S. E. Laux, D. J. DiMaria
IBM T. J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

Coulomb interactions and hot-electron effects in sub-0.1 μm Si MOSFETs

M. V. Fischetti, S. E. Laux, and D. J. DiMaria
IBM Research Division, Thomas J. Watson Research Center
P. O. Box 218, Yorktown Heights, NY 10598, USA

Two-dimensional self-consistent Poisson/Monte Carlo simulations of semiclassical electron transport in sub-100 nm Si devices show that long-range Coulomb interactions dominate the electron dynamics: The fluctuations of the potential in the heavily-doped source and drain contacts penetrate into the channel up to a distance of the order of the Debye length in the channel. In small devices this can be a significant fraction of the channel itself, so that the plasma fluctuations give raise to strongly nonthermal high-energy tails of the electron distribution in the channel itself. Simulations of the hot-electron degradation of the gate insulator emphasize the importance of these Coulomb effects.

I. INTRODUCTION

With the scaling of Si metal-oxide-semiconductor field-effect-transistors (MOSFETs), and the associated reduction of the supply voltage, it was hoped that a significant reduction of hot-electron degradation effects would take place. These hopes have been proven unwarranted both experimentally[1,2] and theoretically[3,4]: Even in relatively 'large' devices, effects caused by electrons at energies above the maximum applied bias have been observed and predicted, such as the presence substrate current at source-to-drain biases below the Si band-gap[1] and of gate currents at biases below the Si-SiO_2 barrier[2]. Short-range electron-electron interactions have been identified as the cause for the presence of carriers at energies above the supplied voltage via a redistribution of kinetic energy among carriers.[3,4]

It is our objective in this paper to show that as devices are scaled even further, at effective channel lengths well below 100 nm, long-range Coulomb interactions become a dominant cause for concern. In the absence of Coulomb interactions, the electron energy distribution at any given location in the channel is known to be truncated at an energy equal to the total voltage drop between the source contact and that location. Above this cutoff energy, the distribution drops sharply with a high-energy tail characterized by the lattice temperature[5], the only temperature scale present in the problem. On the contrary, when accounting for Coulomb interactions, the situation changes: The high-density electron gases in the heavily-doped source and drain contacts are not characterized by the 'conventional' Fermi-Dirac equilibrium distribution typical of non-interacting Fermions. Rather, the mutual Coulomb interactions among electrons and between the electrons and the background dopants result not only in a global lowering of the total energy of the system, but also in the emergence of an extra component of kinetic energy, proportional to the cubic root of the electron density in a semiclassical model. With this extra kinetic energy, δK_C, is associated a new 'temperature scale', T_C, in general much larger than the lattice temperature. In devices small enough for electrons in the channel to be coupled to the high-density electron gases in the contacts, the electron distributions in the channel behave qualitatively 'as expected', but with high-energy tails now characterized by the much larger temperature T_C, as shown in Fig. 1. We shall discuss the origin of these anomalous and unexpected high-energy tails in Sec. II, and their influence on the hot-electron degradation of MOSFETs in Sec III.

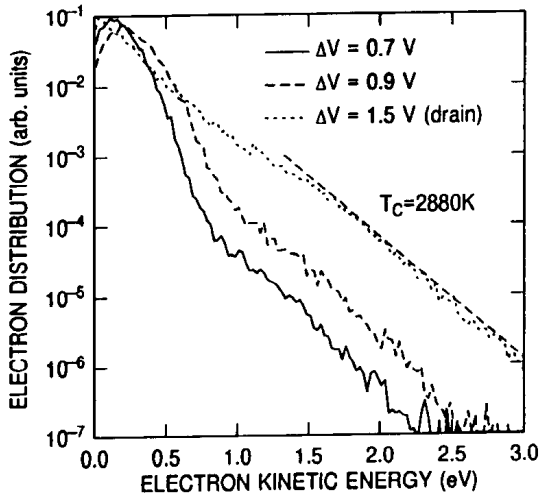

FIG. 1. Calculated electron energy distributions at three locations along the 60 nm-long channel of a Si MOSFET. The device is driven by a drain-to-source bias of 1.5 V, with 1.5 V also applied to the gate across a 2.8 nm-thick oxide. Distributions are shown at two locations along the channel at which the potential drops are 0.7 and 0.9 V, respectively, and at a location deeper inside the drain implant. The peak doping concentration in the drain is about 2.6×10^{20} As atoms/cm^3.

II. ELECTRON GAS AT HIGH DENSITY

The origin of the nonthermal high-energy tails seen in Fig. 1 can be understood by considering the behavior of a homogeneous electron gas.

The ground state of a system of noninteracting Fermions can be obtained by placing at some initial time t_0 all particles in a cubic lattice, assignining them momenta according to the Fermi-Dirac distribution, and letting the system evolve. Since no interparticle interactions perturb the system, the energy distribution will remain constant as a function of time t. When similarly preparing a system of interacting Fermions, it is clear that the initial energy at $t = t_0$ will be the sum of the Fermi energy – as for the noninteracting system – and of the Coulomb potential energy. Various models can be employed to represent the interacting gas and estimate the Coulomb energy. In the Hartree model, both electrons

and background donors are viewed as uniform 'clouds' of opposite charges, so that the system is everywhere electrically neutral and the Coulomb energy identically vanishes. At the next level of approximation, one may view the background dopant as point-like charges embedded in the uniform electron distribution[6]. For this 'modified Hartree' model[7] – actually identical to the complementary model of point-like electrons in a 'jellium' of dopants we employ in our simulations[8] – the Coulomb potential energy is:

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

where ϵ_s is the static, long-wavelength dielectric constant of the semiconductor, n is the electron density, and e is the magnitude of the electron charge. Further levels of approximations account for quantum effects neglected by semiclassical Monte Carlo/Poisson schemes: The Hartree-Fock approximation provides the exchange contribution[9], also negative but about a factor 4 smaller than Eq. (1), while correlation effects can be safely ignored, when using the results of 'exact' calculations[10]. Thus, quantum corrections, while not completely negligible, do not affect too much the semiclassical results. When implementing the three-dimensional semiclassical model just outlined using a two-dimensional Poisson solver, care must be taken to select the correct number of (super)particles and the correct mesh spacing in order to reproduce the result of Eq. (1) and to account correctly for the short-range (*i.e.*, within a single mesh-element) Coulomb interaction, as we have discussed in the past[8]. Finally, note that the shift of the chemical potential of the system induced by the shift δE_C is one of the various components responsible for the narrowing of the bandgap at high doping.[6]

As the system evolves in time, in the presence of random motion – induced by the Fermi kinetic energy, by the Coulomb interactions, and by electron-phonon collisions – the electrons will move away from their initial configuration and the total energy δE_C will redistribute into a potential energy δU_C, and a kinetic energy component δK_C. The virial theorem[11] can be used to estimate the time-average of the kinetic energy δK_C once δE_C is known: For N particles interacting via a central Coulomb potential, $< \delta K_C >= -(1/2) \ < \delta U_C >$, where $< ... >$ denotes the time average, so that:

$$< \delta K_C > = - < \delta E_C > . \qquad (2)$$

A question one may raise at this point is whether this extra kinetic energy and the associated nonthermal tails of the electron energy distributions are at all consistent with what we have to expect at thermodynamic equilibrium. The answer is simple: These non-thermal tails are the result of enforcing global equilibrium at the lattice temperature, overlaying an equilibrium Fermi-Dirac distribution on a spatially and temporally fluctuating potential. Indeed the random motion of the electrons is obviously associated with fluctuations of the potential (plasma oscillations of various wavelengths). We can describe them by a function $W(\delta V)$, expressing the probability of finding at any given time and given location a value V for the potential, distant by an amount $\delta V = V - V_0$ from its space and time averaged value V_0. If we select as initial and/or boundary condition (as reflected by the 'Ohmic contacts' employed by the simulation) a Fermi-Dirac distribution $f_{FD}(\epsilon)$, where $\epsilon = K + V$ is the *total* electron energy, K and V being the kinetic and

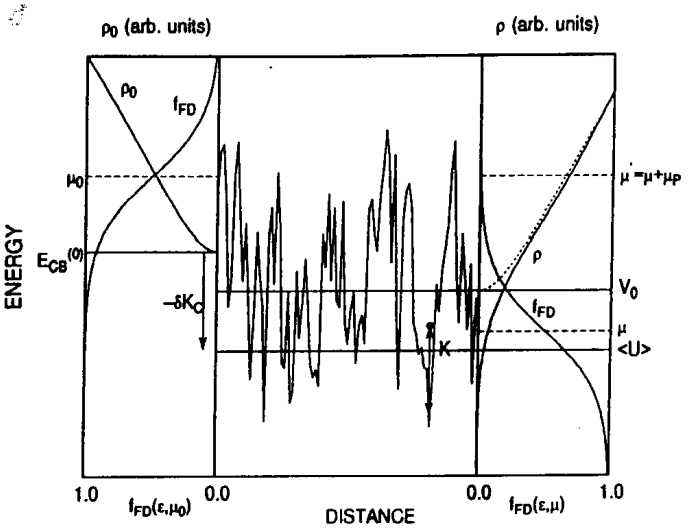

FIG. 2. Schematic diagram illustrating graphically the origin of the Coulomb kinetic energy δK_C and of the associated high-energy tails. If the gas is assumed to reach a global equilibrium population $f_{FD}(\epsilon)$ in total energy $\epsilon = K + V$, shown by the curve at right, the presence of potential fluctuations cause electrons even at moderate total energies to populate regions of high kinetic energy K. The dot represents an electron of relatively small total energy which, however, will contribute to the large kinetic-energy tails. Note that the 'single particle' Fermi energy μ is related to the equilibrium chemical potential of the reservoirs μ_0 via $\mu_0 = \mu + \mu_P$, where μ_P originates from the collective plasma oscillations. Note that the extra kinetic energy δK_C is compensated by a shift of average potential energy, $< U >$, as required by conservation of total energy.

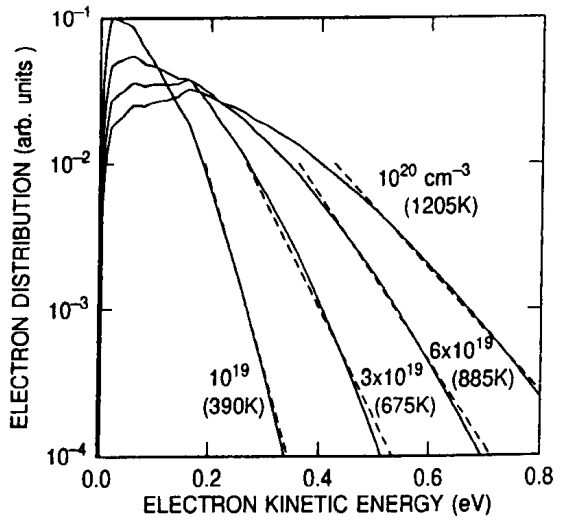

FIG. 3. Kinetic energy distributions for a homogeneous electron gas at zero field at the indicated electron densities, calculated from potential fluctutations, Eq. (3). The dashed lines are exponentials at the temperatures shown in parenthesis.

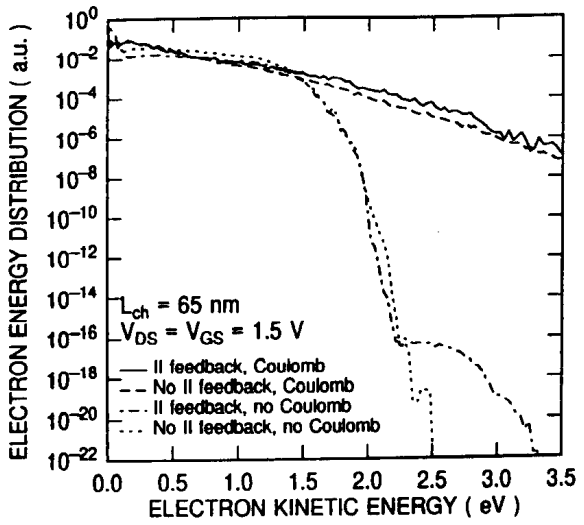

FIG. 4. Calculated electron energy distribution integrated over a region close to the channel/drain junction of a 60 nm-long channel Si MOSFET. The 'colder' distributions have been obtained from non self-consistent simulations neglecting Coulomb interactions, while the 'hotter' distributions show the effect of the (mainly long-range) Coulomb processes. The small effect of the impact-ionization feedback is also shown.

potential energy, respectively, the kinetic energy distribution, $n(K)$, will be the equilibrium function $f_{FD}(\epsilon)$ weighted by the potential fluctuation distribution $W(\delta V)$:

$$n(K) = \rho_0(K) \int_{-\infty}^{\infty} d\epsilon \; f_{FD}(\epsilon) \; W(\epsilon - K - V_0) \; , \quad (3)$$

where $\rho_0(K)$ is the DOS at kinetic energy K, here taken from empirical pseudopotential calculations for Si, $f_{FD}(\epsilon) = \{1 + \exp[\beta(\epsilon - \mu)]\}^{-1}$, ($\beta$ being the inverse temperature) and the global 'single-particle' Fermi potential μ must be determined from charge neutrality. The physical meaning of Eq. (3) is quite simple, as illustrated in Fig. 2: A constant single-particle Fermi level μ implies that many particles will sit in the troughs of the potential fluctuations, none at the peaks. As a result, the particles, while having *total* energy consistent with global equilibrium at the lattice temperature, will have a high kinetic energy, consistent with the high tails observed in the kinetic energy distributions. Note also that at a finite temperature T the total chemical potential of the system, μ', is given by[12]:

$$\mu' = \frac{\partial E}{\partial n} + T \frac{\partial S}{\partial n} \; , \quad (4)$$

where E and S are the total energy and entropy density of the gas. They depend on the electron density n both via the single-particle Fermi potential μ, as for non-interacting systems, and via the n-dependent potential fluctuation distribution W. This last contribution, here denoted by μ_P, physically represents the energetic cost of adding one particle to the gas while exciting collective (plasma) modes. Therefore $\mu' = \mu + \mu_P$, not μ, must equal the unperturbed Fermi level in the contact, μ_0, when the gas is in equilibrium with the reservoir. Figure 3 shows the kinetic energy distributions obtained from this procedure. Particularly noticeable are the high-energy tails described by the same large effective temperature observed in the 'nonthermal' tails obtained from MC simulations.

III. HOT ELECTRON DEGRADATION

The potential fluctuations $W(\delta V)$ just discussed are obviously present in the heavily-doped drain and source contacts of MOSFETs. They will penetrate the channel of the device up to a distance of the order of a few Debye lengths. In short devices, especially near pinch-off (*i.e.*, depleted channels with long Debye lengths), these long-range collective Coulomb effects will influence most of the channel and contribute to the build-up of the same high-energy tails in the channel itself.

In order to assess the role played by these effects in the hot-electron degradation of devices, we have simulated an n-channel Si MOSFET with an effective channel length of about 60 nm, with shallow source and drain pockets doped to about 3×10^{20} cm^{-3}, halo implantation and retrograde doping in the p-well. We have employed our simulation program DAMOCLES[8] including dynamically-screened short-range particle-particle interactions, as described in Ref. 13. For comparison, we have employed four different simulation conditions: 1. 'Control' simulations, employing a Monte Carlo algorithm to handle both electron and hole transport, and accounting fully for Coulomb effects via the self-consistent Monte Carlo/Poisson iterations and by including short-range interparticle interactions. The use of a Monte Carlo model to treat hole transport allows us to account also for the 'impact-ionization feedback' process[14]. 2. A similar set of self-consistent simulations, but now treating hole transport with a simpler drift-diffusion approximation *and* suppressing hole-initiated impact ionization. This allows us to estimate directly the role of the impact-ionization feedback process. 3. Non self-consistent simulations using a Monte Carlo scheme to treat electron and hole transport, as in 1. above, but now employing a 'frozen' field (obtained from a suitable time average of the results from the self-consistent runs), and suppressing the short-range Coulomb processes. 4. Finally, non self-consistent runs with drift-diffusion holes.

Figure 4 shows the electron energy distributions we have obtained from the four simulation conditions just described at a gate-source and drain-source bias of 1.5 V and at a lattice temperature of 300 K. The distributions have been gathered over a region surrounding the surface channel/drain junction. The non self-consistent results show the expected thermal fall-off at an energy given roughly by the applied bias. The effect of the impact-ionization feedback is clearly seen in this case. The introduction of the Coulomb interactions results in the expected build-up of high-energy tails at the temperature $T_C \approx$ 2900K corresponding to the total kinetic energy (Fermi + Coulomb, $K_F + \delta K_C$) at the large electron density ($\approx 2.6 \times 10^{20}$ cm-3) present in the drain region. Results of simulations performed accounting for the long-range Coulomb interactions (via the Poisson/Monte

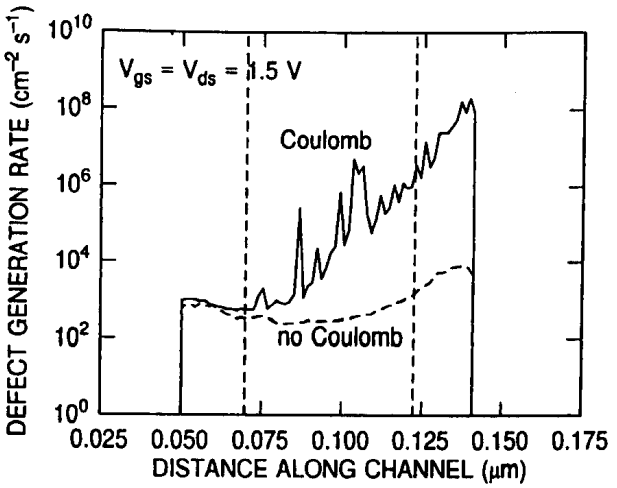

FIG. 5. Generation rate of defects at the Si-SiO_2 interaface calculated using the data of Ref. 17. The results obtained accounting for (long- and short-range) Coulomb interactions are compared with those obtained ignoring Coulomb effects.

Carlo self-consistency), but suppressing the short-range interparticle collisions, show that the latter ones are not the dominant effect, unlike what is found in devices with longer (>150nm) channels. In any event, Coulomb interactions, short- and long-range, dominate over the impact-ionization feedback in enhancing the high-energy tails of the electron distributions[15]. The short-range interparticle collisions have already been shown to be more important than impact-ionization feedback[16]. Here we find that for very small devices even more important are long-range collective modes excited in the drain and penetrating well into the channel.

To further assess the importance of Coulomb effects, we have employed the degradation data of Stathis and DiMaria[17] to estimate the generation of defects in the oxide: During the Monte Carlo simulation, we have gathered the energy distribution of electrons hitting the Si-SiO_2 interface and of those emerging into the gate after tunneling across the gate insulator, as described in Ref. 13. The probability of generating defects at either interface was estimated using the model of Ref. 17 expressing the generation rate as a function of electron kinetic energy at either interface. An example of the results is shown in Fig. 5: The rate at which interface states are generated along the Si-SiO_2 interface is shown to be significantly higher, particularly at the drain-side of the channel, when accounting for Coulomb effects.

IV. CONCLUSIONS

We have shown that in ultra-small Si MOSFETs the behavior of the hot electrons in the channel can be significantly affected by the interaction of these electrons with the high-density electron gases in the source and drain contacts. Collective excitations in these heavily-doped regions can penetrate a large fraction of the channel, thus building highly nonthermal high-energy tails in the electron energy distributions. These effects dominate over single-particle (short-range) Coulomb interactions (which dominate in longer devices), and also over additional high-energy processes, such as the impact-ionization feedback mechanism[14].

[1] L. Machanda, R. H. Storz, R. H. Yau, K. F. Lee, and E. H. Westerwick, IEDM Tech. Dig., 994 (1992).

[2] S. Tam, F.-C. Hsu, C. Hu, R. S. Muller, and P. K. Ho, Electr. Dev. Lett. **ED-4**, 249 (1983).

[3] C. C. C. Leung and . A. Child, Appl. Phys. Lett. **66**, 162 (1995).

[4] M. V. Fischetti and S. E. Laux, IEDM Tech. Dig., 305 (1995).

[5] First noticed by A. Lacaita, Appl. Phys. Lett. **59**, 1623 (1991), the origin of these 'high-energy' thermal tails has been studied by A. Abramo, C. Fiegna, and F. Venturi, IEDM Tech. Dig., 301 (1995) and by A. Pacelli and A. L. Lacaita, Semicond. Sci. Technol. **11**, 1642 (1996).

[6] G. D. Mahan, J. Appl. Phys. **51**, 2634 (1980) has shown that at large electron densities (> 10^{19} cm^{-3}) the more conventional way of looking at the problem – each electron 'lumped' around each impurity so to screen it according to the Debye-Hückel model, yielding the Yukawa-like impurity potential – is inconsistent and it is actually a worse approximation than the assumption of a uniform electron distribution.

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

[8] See, for instance, M. V. Fischetti and S. E. Laux, Phys. Rev. B **38**, 9721 (1988). This references discusses the problem of double-counting long-range and short-range Coulomb interactions and of the plasma oscillations. More numerical details are given in S. E. Laux and M. V. Fischetti, in *Monte Carlo Device Simulation: Full Band and Beyond*, Karl Hess ed. (Kluwer, Boston, Massachusetts, 1991).

[9] C. Haas, Phys. Rev. **125**, 1965 (1962).

[10] M. Gell-Mann and K. Brueckner, Phys. Rev. **106**, 364 (1957); P. Nozières and D. Pines, Phys. Rev. **111**, 442 (1958).

[11] H. Goldstein, *Classical Mechanics*, 2nd edition, (Addison-Wesley, Reading, Massachusetts, 1980), Chp. 3.

[12] R. Kubo, *Statistical Mechanics*, (North-Holland/American Elsevier, New York, 1965).

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

[14] J. D. Bude, Sym. on VLSI Tech. Tech. Dig., 101 (1995).

[15] Another reasons why the impact-ionization feedback process plays a decreasing role in smaller devices lies in the fact that scaled-down devices must employ shallower source and drain junctions. Holes generated by impact ionization initiated by channel electrons will in turn ionize deeper in the substrate. In small devices, these secondary ionization process will occur mainly in the low-field (or even retrogade field, for retrogade doping) beyond the drain/body junction. Many of the generated (feedback) electrons will be more likely to drift within the substrate, rather than be swept back towards the Si-SiO_2 interface.

[16] J. D. Bude, T. Iizuka, and Y. Kamakura, IEDM Tech. Dig., 865 (1996).

[17] J. Stathis and D. J. DiMaria, IEDM Tech. Dig., 167 (1998).