Simulation of Electron Transport in High-Mobility MOSFETs: Density of States Bottleneck and Source Starvation

M.V. Fischetti, L. Wang†, B. Yu†, C. Sachs, P. M. Asbeck†, Y. Taur†, and M. Rodwell*
Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003, USA
† Department of Electrical and Computer Engineering, University of California, San Diego, CA 92092, USA
* Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, USA

Abstract

20 nm InGaAs-based MOSFETs are studied using dc and transient Monte Carlo simulations and self-consistent Schrödinger-Poisson solutions accounting for nonparabolic corrections. The latter simulations show that nonparabolicity can boost the carrier concentration in the InGaAs channel by up to 35% with respect to calculations based on parabolic models, while Monte Carlo simulations show that an optimization of the source, channel, and source/channel regions can significantly improve the performance of the devices. This optimization overcomes a problem that results from the low density-of-states (DOS) in materials with low effective mass, which appears in the quasi-ballistic limit: The inability of the source region to sustain a large flow of carriers in 'longitudinal' velocity states in the channel ('source starvation'), unless the momentum relaxation rate and/or the doping density in the source are sufficiently large.

Introduction

The search for materials allowing CMOS technology to sustain Moore's law beyond 'conventional' silicon FETs focusses on high-mobility materials because of theoretical results linking the low-field mobility to the on-state performance (as measured by on-current and transconductance)[1]. As devices shrink to the length-scale at which quasi-ballistic transport is expected to occur, an additional motivation stems from the fact that a small effective mass, m^*, translates into a larger ballistic velocity v at a given kinetic energy E, $v \sim (2E/m^*)^{1/2}$, and into a reduced scattering rate, $1/\tau$, since $1/\tau \propto m^{*3/2}$ assuming that the scattering rate scales with the density of states (DOS)[2]. Because the maximum current the device can sustain is proportional to the product nv, where n is the density of carriers, this simple picture is accurate only in the limit of the DOS capacitance C_{DOS} being much larger than the insulator capacitance C_i, where n is independent of m^*. In the opposite limit of small devices with extremely thin insulators, i.e., $C_i \gg C_{DOS}$, n scales as DOS $\sim m^{*3/2}$ in bulk and $\sim m^*$ in 2D electron gases, so that the current scales as $m^{*1/2}$ to m^*. Thus in the ballistic limit, a *larger* effective mass may be desirable[3, 4]. The DOS limit appears because the low DOS of small-mass materials implies that a large swing of the Fermi level is required to change substantially the carrier density in the channel. In turn, this requires a large swing of gate voltage to change the drain current, and the resulting transconductance is depressed. The effect has become commonly known as the 'DOS bottleneck'. In representative devices, this effect is mitigated by the fact that the strong non-parabolicity exhibited by small effective mass materials enhances the DOS and thus the charge density. Therefore, detailed calculation is required in order to compare FETs based on different material systems.

Here we show an additional effect which is important in the design of MOSFETs with high mobility channels related to low DOS. As transport approaches the ballistic regime, calculations of the gate capacitance which assume an equilibrium distribution of carriers close to the source/channel junction will necessarily result in optimistic overestimations: While 'streaming' ballistically, carriers populate mainly states of wavevector $\mathbf{k}$ (or velocity) aligned along the transport direction (we shall call these states 'longitudinal $\mathbf{k}$ states'). Momentum randomizing collisions are required to replenish 'transverse $\mathbf{k}$ states'. In absence of collisions, the density drops significantly when compared to the equilibrium value. 'Fast', light-mass materials approach the ballistic regime at lengths longer than heavier-mass semiconductors.

In this paper, we study theoretically the effect of small-mass channel-materials in MOSFETs at the 20 nm gate-length. We show that conduction band nonparabolicity significantly mitigates the reduction of inversion capacitance in representative designs, and sheet carrier densities as high as 10^{13} cm^{-2} are predicted with InGaAs channel MOSFETs at gate voltage Vg $\sim$ 1V. We show that for representative designs the expected output current can be higher than that of Si. We also show the importance of the DOS bottleneck as transport approaches the quasi-ballistic limit: The strong flux of carriers exiting the source may cause a starvation of carriers occupying longitudinal $\mathbf{k}$-states, thus rendering the source unable to supply enough carriers and sustain a high on-current. We also show that an improved source/channel design may reduce these effects.

Simulation Method and Device Structure

In our study we have considered 20 nm gate-length $In_xGa_{1-x}As$-channel (x=0.53) MOSFETs with a 3.7 nm-thick HfO2 gate insulator on an InAlAs insulating barrier and p-doped InGaAs ground planes. The nominal devices consist of a 2.5 nm-thick InGaAs/2.5nm InP channel and source/drain regions doped to 2×10^{19} donors/cm^3, as shown schematically in Fig. 1.

1-4244-0439-X/07/$25.00 © 2007 IEEE

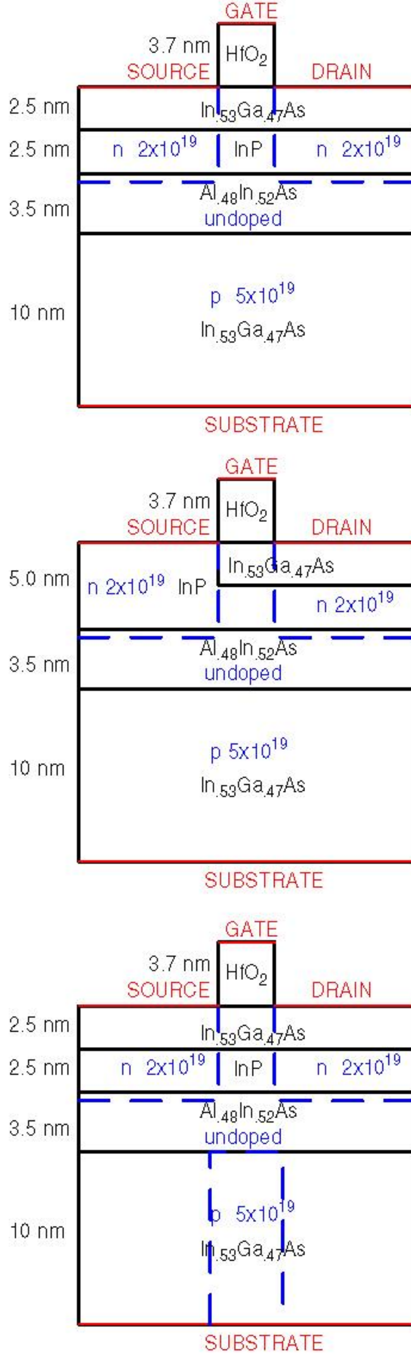

Figure 1: Schematic diagram illustrating the geometry of the simulated devices. Top frame: Original device. Middle frame: InP launcher. Bottom frame: Ground-plane doping only under the gate.

Calculations of sheet carrier density (N_s) vs. voltage for a long channel device have been performed using a Schrödinger-Poisson solver which takes into account non-parabolicity. Assuming the non-parabolicity can be expressed as $\mathrm{T}(1+\alpha\mathrm{T}) = \frac{\hbar^2}{2m^*}(-\frac{\partial^2}{\partial z^2}) + \frac{\hbar^2}{2m^*}K_{xy}^2$, where α is the non-parabolicity parameter associated with the material, an iterative version of Schrödinger equation can be derived. Figure 2 shows the equilibrium N_s vs. V_g. N_s reaches 1×10^{13} cm^{-2} larger by a factor 1.35 than results obtained by neglecting nonparabolicity.

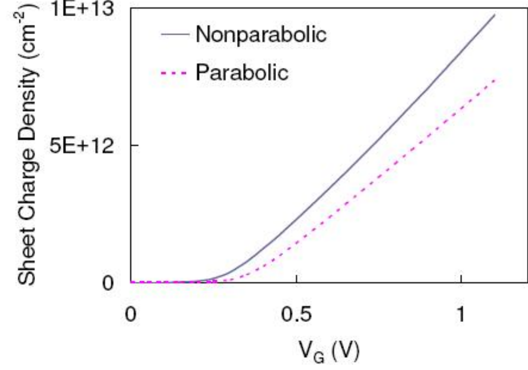

Figure 2: Equilibrium sheet carrier density vs. gate voltage for long-channel devices computed using a Schrödinger-Poisson solver including nonparabolic effects.

I-V characteristics have been simulated using Monte-Carlo methods[5] accounting for scattering with high-κ-related interfacial optical phonons[6]. The relatively low I_{DS} obtained with this device design (which we label device type '1') have prompted us to consider several variations of the 'nominal' MOSFET. We have considered devices type '2', with an InP launcher in the source region (Fig 1 middle, attempting to boost the injection velocity[7]); '3', with a lightly-doped source extension (Fig.1 bottom, to ameliorate DOS-bottleneck effects); '4', with the p^+-doped ground-plane only under the gate (to minimize source-depletion); '5', with more heavily doped (5×10^{19} cm^{-3}) S/D regions; '6', raised source/drain (S/D) regions (to enhance electron supply in the source); and, finally, of type '7', with a thicker (5 nm) InGaAs channel. A nominal device but with a Si channel (2.5 nm thick) and HfO_2 top and bottom barriers has also been considered for comparison.

Monte Carlo results and discussion

Figure 3 illustrates the turn-on characteristics of the various devices at V_{DS} = 1.0 V. Figure 4 summarizes the results, showing that heavily-doped S/D, raised S/D, and thicker channels improve dramatically the transconductance. All of these design variations allow electrons in the source to scatter and effectively replenish longitudinal $\mathbf{k}$-states which are injected into the channel in ballistic-transport conditions (as illustrated schematically in Fig. 5).

This explanation suggests that it can be erroneous to as-

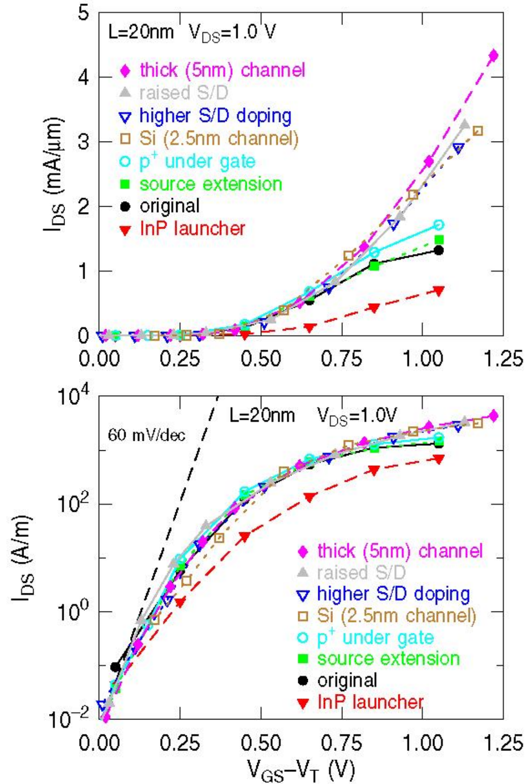

Figure 3: Calculated turn-on characteristics at V_{DS}=1.0 V. (top frame: linear; bottom frame: semi-log, curves shifted to account for the different threshold voltage, V_T, in the various device designs) of the simulated devices. Note, from the sub-threshold swing shown, the good short-channel behavior exhibited by most of the device designs considered here.

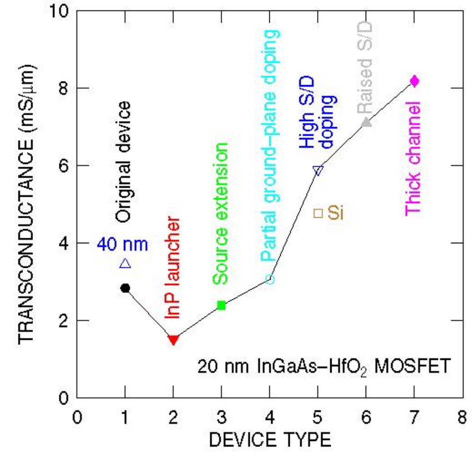

Figure 4: Calculated transconductance of the various device designs considered. The transconductance figures of a 40 nm-long nominal device and of a Si-channel device are also shown.

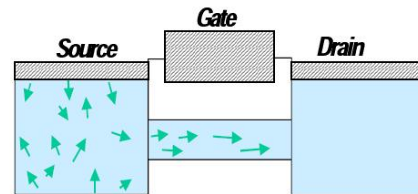

Figure 5: Structure of a MOSFET showing schematically the distribution of electron velocities. Scattering is needed in order to replenish carriers entering the channel.

sume an equilibrium carrier distribution, not only near the source-channel junction or in the channel, but also a few diffusion-lengths inside the heavily-doped source region. This effect may be called 'source starvation': Within or near the ballistic regime, electrons are carried away so efficiently along the channel that the source region becomes 'starved' of carriers populating longitudinal **k** states. Unless the momentum relaxation time in the source becomes sufficiently short, there will not be enough carriers which can be injected into the channel and contribute to the drain current: The source becomes 'starved' of electrons in 'longitudinal **k**-states'. A higher doping in the source supplies more carriers which can be scattered and redirected into longitudinal **k** states. It also reduces the momentum relaxation time, as there are more impurities to which electrons can transfer momentum. This is shown in Fig 6, illustrating the current (ballistic in the channel) as a function of the momentum relaxation time in the source (due mainly to electron-impurity collisions) expected simply from geometrical considerations and classical kinetic theory. Raised S/D increase the 'acceptance' cone, effectively enlarging the volume over which scattering can redirect carriers into the longitudinal **k** states. Finally, a thicker channel also helps in widening the acceptance cone.

Note also how increasing the injection velocity (as found in the InP-launcher device, devices of type '2') worsens the situation, by depleting more severely the injecting contact. Indeed, as the ballistic limit is approached and/or the injection velocity increases, this "source starvation" effect becomes progressively more important. Figure 4 also shows the maximum g_m simulated for the optimized InGaAs device. Also shown for comparison is the maximum g_m (at the same gate overdrive) of a 2.5 nm-thick Si-channel device with HfO_2 bottom barrier, 2×10^{20} cm^{-3} S/D doping, but otherwise similar to the nominal InGaAs device.

Finally, while the dc transconductance may be regarded as a good figure of merit to gauge device performance, the simulation of transient response of the device may better correlate with the actual speed of the device in a circuit. Here we have followed the procedure employed by Frank *et al.*[8], employing the simple circuit illustrated in the inset of Fig. 7 to estimate the response of a loaded

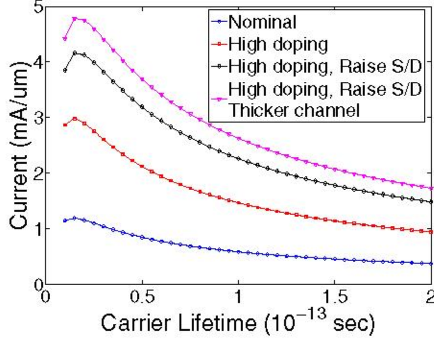

Figure 6: On-current calculated for purely ballistic transport assuming an injection velocity independent of doping, but varying the momentum-randomization time in the source, assuming that geometrical effects control the rate at which electrons are injected into the channel. As scattering in the source becomes more frequent, more electrons are scattered into longitudinal **k**-states and can contribute to the current. Nominal device has 2×10^{19} cm^{-3} S/D doping; high doping refers to S/D doping as 5×10^{19} cm^{-3}.

inverter. Figure 7 shows the rising and falling waveforms for the type '7' device. The average delay time extracted from these simulations is 0.208 ps. Similar simulations yield an average delay time of 0.329 ps for the type '4' device. Comparing these delay times with the respective on-current and transconductance figures (1.7 mA/μm and 3.0 mS/μm for device '4' *vs.* 4.3 mA/μm and 8.1 mS/μm for device '7') we see that the dc characteristics overestimate the differences between these two types of devices. They both appear to outperform an equivalent Si device (with a 2.5 nm thick channel with HfO_2 gate insulator and buried insulator on the ground plane) which exhibits an average delay of 0.345 ps. The best III-V device, in conclusion, may outperform the Si device by roughly 40%, thanks to reduced capacitance.

Results and conclusions

In conclusion, we have simulated InGaAs-based MOSFETs at the 20 nm scale and have shown that optimization of the design of the source, channel, and channel/source regions can minimize the damaging effects of the DOS bottleneck. In so doing we have uncovered a new flavor of the DOS-bottleneck itself, namely the inability of a low-doped source to sustain the large flow of (quasi)ballistic carrier in the channel. This effect, which we have termed 'source starvation', manifests itself as devices are shrunk towards the ballistic regime. We have shown that nonparabolic corrections are required to obtain realistic values for the maximum carrier concentration achievable in the channel and also contribute in minimizing the DOS bottleneck. Finally, it appears the InGaAs-based devices can, ideally, outperform equivalent Si-based devices, but careful optimization must be performed. This work has been supported by the SRC.

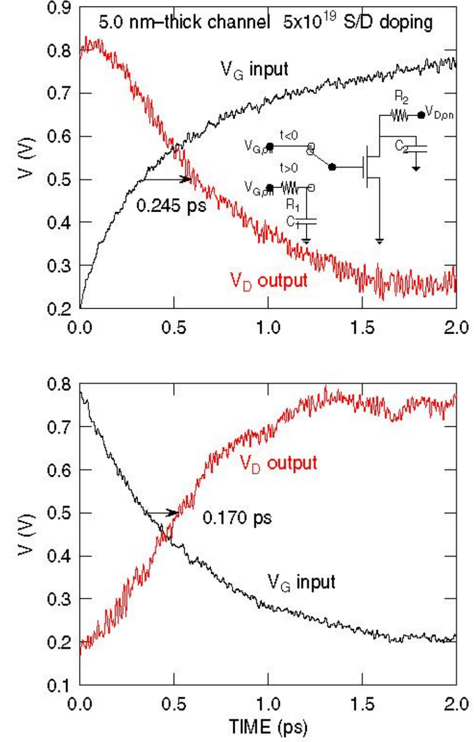

Figure 7: Simulated transient response for the device type '7' of the text (5 nm thick channel, S/D regions doped to 5×10^{19} cm^{-3}): input rising/output falling (top frame), and input falling/output rising (bottom frame). The inset in the top frame shows the circuit used in the simulation. The values for the parameters R_1, R_2, C_1, and C_2 were chosen so that the output response of the bottom frame emulates the input of the top frame.

References

[1] M. Lundstrom, "On the mobility versus drain current relation for nanoscale MOSFETs", *Electron Dev. Lett.*, vol. 22, no. 6, pp. 293–295, 2001.

[2] H. Shichijo and K. Hess, "Band-structure-dependent transport and impact ionization in GaAs", *Phys. Rev. B* vol. 23, no. 8, pp. 4197-4207 (1981).

[3] P. M. Solomon and S. E. Laux, 'The ballistic FET: design, capacitance and speed limit", *IEDM*, pp. 95–98, 2001.

[4] M. V. Fischetti and S. E. Laux, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 39, no. 3, pp. 650–660, 1991.

[5] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", *Phys. Rev. B*, vol. 38, no. 14, pp. 9721–9745, 1988.

[6] M. V. Fischetti, T. O'Regan, S. Narayanan, C. Sachs, S. Jin, J. Kim, and Y. Zhang, "Theoretical Study of Some Physical Aspects of Electronic Transport in nMOSFETs at the 10 nm Gate-Length", *IEEE Trans. Elec. Dev.*, vol. 54, no. 9, pp. 2116–2136, 2007.

[7] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 49, no. 1, pp. 133–141, 2002.

[8] D. J. Frank, S. E. Laux and M. V. Fischetti, "Monte Carlo Simulation of a 30 nm Dual-Gate MOSFET: How Short Can Si Go?", *IEDM*, pp. 553–556, 1992.