Electron Velocity in sub-50-nm Channel MOSFETs

Dimitri A. Antoniadis, Ihsan J. Djomehri, and Anthony Lochtefeld
Microsystems Technology Laboratories
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract

Inverse modeling of state-of-the-art NMOSFETs is used to investigate electron transport models and in particular to extract the effective velocity of electron injection from source to channel. It is found that this velocity is less than 50% of the maximum possible velocity, i.e. the thermal velocity of electrons in the source. Based on the Landauer formulation, as adapted by Lundstrom to silicon MOSFETs this indicates that modern NMOSFETs are quite far from their ballistic transport limit and therefore their current is still limited by momentum scattering as manifested in the electron mobility. Investigation of mobility in those transistors reveals that it is reduced with channel length, most likely due to Coulomb scattering by the ionized dopant atoms in the source and drain halos that are necessary for well-tempered ultra-short-channel MOSFETs, and possibly remote scattering by the source dopants.

1 Introduction

Investigation of electron transport models by comparison of simulated to experimental devices requires accurate knowledge of the experimental device structure and doping configuration. To-date there are no direct doping characterization techniques for deeply scaled MOSFETs, and therefore one has to resort to indirect characterization methods. Inverse modeling (IM) based on 2-D device simulation has proven to be a very powerful such method. While at first sight this may appear as a vicious circle, in fact using electrical measurements in a regime where device operation is very sensitive to electrostatic potential and not very sensitive to transport models resolves the dilemma. Weak-inversion is such a regime for I-V characteristics [1], and C-V also has no dependence on carrier transport.

In this work, combination of weak-inversion I-V and C-V measurements of devices from two different technologies, that here will be referred to by their physical oxide thickness, t_{ox}, 3.3 nm and 1.5 nm. Accurate doping S/D extension and channel profiles are obtained by this method and are then used to investigate electron transport in these devices.

IM-extracted values of the source series resistance, R_S, are used in a new method to extract electron velocity nearest to the source-channel injection point than any other method, and this velocity is compared with the theoretical values of thermal injection velocity in order to assess how close to the ballistic limit modern NMOSFETs operate. Further, IM-extracted values of effective channel length, L_{eff}, are used along with values of R_S to extract low-longitudinal-field mobility in extremely short channel devices.

Given that all standard methods for L_{eff} and R_{SD} characterization fail in the case of modern sub-100-nm devices, IM provides new insight into mobility at short channels.

2 Inverse Modeling for MOSFET Chararterization

Figure 1 shows the inverse modeling flow diagram. Because the subsequent log(I)-V data is sensitive to the gate stack, the first step is to accurately account for it by extracting the oxide thickness and poly doping from the accumulation and inversion regimes of the C_{gg}-V curve, respectively. Quantum Mechanical effects are captured via an approximate band gap shift [2]. Now, 1-D IM is performed to extract the channel doping on a long-channel device. No process simulation or SIMS were availabe here, but could be used to help construct a good initial guess to the 2-D profiles of the short-channel devices. Next, detailed TEMs were necessary to completely characterize the gate stack and spacer dielectric structures to capture the effect of the fringing capacitances. Finally, the combined log(I)-V and C-V IM approach alternated between a few iterations of each data set using a standard non-linear least-squares optimization algorithm. The effect of small variations in the 2-D doping parameterization on the RMS error are used to obtain a converging solution to the parameterization (which consists of two 2-D gaussians for S/D extensions and halos, plus the long-channel profile).

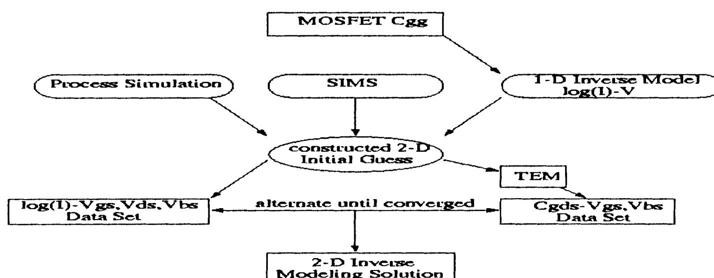

Figure 1. Flowchart delineating ideal comprehensive log(I)-V & C-V IM methodology

Once reliable doping profiles are obtained for several devices of different gate length in each technology, it is possible to proceed with the investigation of electron transport. First, electron mobility vs. transverse electric field, the so-called universal mobility is calibrated for each technology using long channel devices and the split-C-V method for extraction of μ_{eff} vs. E_{eff}. Note that since different nitridation processes were used in the fabrication of the two technologies, the universal effective mobility curve is only universal within each technology. The generalized low-field mobility model [3] from MEDICI is used to give the minimum of the bulk mobility, assumed to be for minority carriers in bulk, and the universal mobility. The universal vertical field dependence is assumed to have an inverse cube root term and an inverse to a higher power which is calibrated along with the magnitudes of these terms.

With the μ_{eff} vs. E_{eff} model in hand, next the source/drain series resistance R_{SD} is

extracted by fitting simulated to measured I_D-V_{GS} at low V_{DS} (typically 50 mV). Note that there are several components of R_{SD}. The V_{GS}-dependent component of the source/drain extension is included in the simulation via the the IM-extracted S/D doping profile, but the deep-S/D and silicide contact resistance are lumped into a single resistance value. For the two technologies under investigation we find the R_{SD} is 220 and 245 $\Omega-\mu$ m, for t_{ox} 1.5 and 3.3 nm at V_{GS} 1.0 and 1.8 V, respectively.

Then, the high longitudinal field transport model is investigated. First, for moderately short-channel devices (L_{eff} > 100 nm), the fitting parameters beta and v_{sat} in the Caughey-Thomas expression for the EB mobilitly model [4] are obtained by matching strong inversion I-V at high V_{DS}. Next, a single value for the carrier relaxation time is found that matches the strong inversion data for the entire family of short devices.

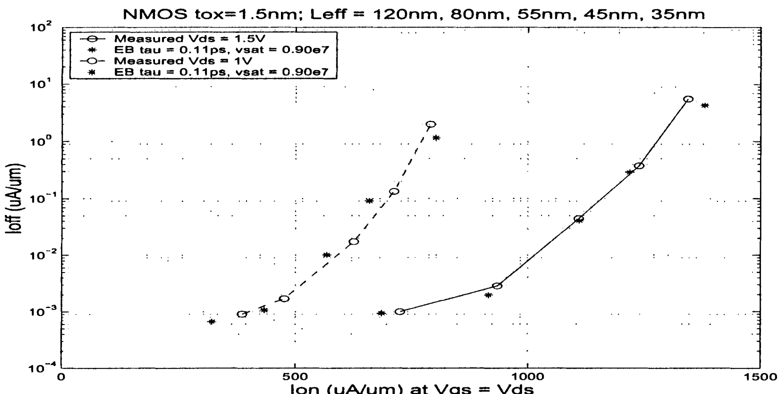

Figure 2. Energy balance fit to experimental I_{on} vs. I_{off} curves at two V_{DD}s

Figure 2 shows the resulting I_{Dsat} vs. I_{off} (V_{DS}=V_{GS}=V_{DD}) for the t_{ox} = 1.5 nm technology. For all devices the high longitudinal field transport model parameters are the same, while the μ_{eff} vs. E_{eff} is unique to each technology. While the Energy Balance model is a simplification it is gratifying to observe that the values for v_{sat} and low-T_e relaxation time constant are both unique and consistent with experimental and theoretical values, respectively [5, 6].

3 Source-Channel Injection Electron Velocity

It has been shown that electrons in a nMOSFET channel can exceed significantly v_{sat} for isotropic field regions (approx. 10^7 cm/s) [7]. The ultimate limit of drain current is, rather, thought to be determined by the thermal injection velocity, v_{inj}, from the source accumulation region into the channel [8]. This can be stated as

$$I_{on}/W = v_{inj}Q_i(x_o)(T/(2-T)) \quad \text{Eq. (1)}$$

Where $Q_I(x_o)$ is the normalized inversion charge at the conduction-band peak at $x=x_o$ at the source side of the channel. Eq. 1 then defines the *effective source-channel injection velocity* at x_o as $v_{eff}=v_{inj}T/(2-T)$, where T is the transmission coefficient at x_o which depends on momentum scattering rate, and barrier width. T=1 and therefore $v_{eff}=v_{inj}$, represents the fully ballistic limit where there is no backscattering of electrons into the source. Clearly, v_{eff} is purely diffusive because the longitudinal electric field is 0 at x_o, and is of great significance in understanding the ultimate limit of MOSFET performance but is also very difficult to measure experimentally.

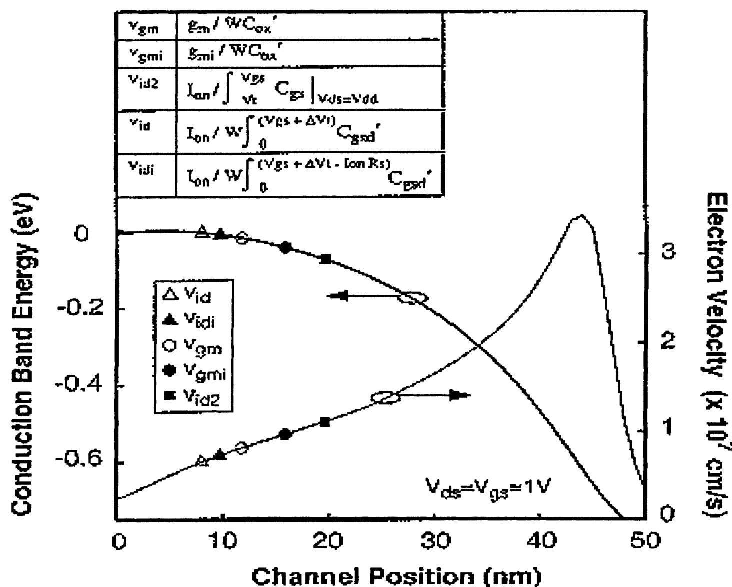

Figure 3. Conduction band and carrier velocity versus position in bulk L_{eff} = 50 nm NMOS.

Several methods have been reported for measuring near-source carrier channel velocity experimentally [7, 9, 10]. In order to evaluate these experimental velocities relative to the well-defined v_{eff} we have simulated measurement techniques and plotted the extracted electron velocity for each on the simulated velocity vs. channel position curve, for a modern nMOSFET similar to the shortest t_{ox}=1.5 nm inverse-modeled ones. Figure 3 shows the results, and the inset summarizes the experimental velocity extraction methods. Two observations can be made: 1) None of the techniques yields a velocity at x_o but the most recent method [10], with or without correction for R_S yields velocities closest to x_o. 2) Both v_{gm} and v_{gmi}, the latter from *intrinsic* g_m corrected for R_S [11], while very easy to obtain represent velocities rather far downstream from x_o.

It has been pointed out [12] that velocity vs. DIBL is a characteristic merit curve unique to each technology, similar to I_{Dsat} vs. I_{off} ($V_{DS}=V_{GS}=V_{DD}$); in both cases, channel length is an implicit variable. Independent of technology, the highest perfor-

mance well tempered MOSFETs generally exhibit DIBL less than 100-150 mV/V, and therefore is best to compare devices from different technologies at constant DIBL. We arbitrarily define v_{idi} at DIBL=100 mV/V to represent the upper limit of v_{eff} for a given technology. Table 1 shows the values of this v_{eff} for the two experimental technologies as well as for a Monte Carlo simulated 25 nm nMOSFET [13]. As can be seen the value of v_{eff} and therefore the transmission coefficient is well below the ballistic limit, and appears to be decreasing with scaling. Since it is desirable to increase v_{eff} for commensurate device performance with scaling we examine next why it decreases instead.

Table 1. v_{eff} for: Monte Carlo; Tech. A = t_{ox} 1.5 nm, V_{DD} 1 V; Tech. B = t_{ox} 3.3 nm V_{DD} 1.8 V

	25 nm M.C.	Tech. A	Tech. B
v_{eff} (cm/s)	$6.1\text{-}7.0\times10^6$	6.7×10^6	7.9×10^6
v_{eff} / v_{inj}	0.35-0.40	0.39	0.49
T	0.52-0.57	0.56	0.66

← Decreasing L_{eff}, T_{ox}
← Increasing v_{inj}

4 Electron Mobility in Ultra-Short-Channel MOSFETs

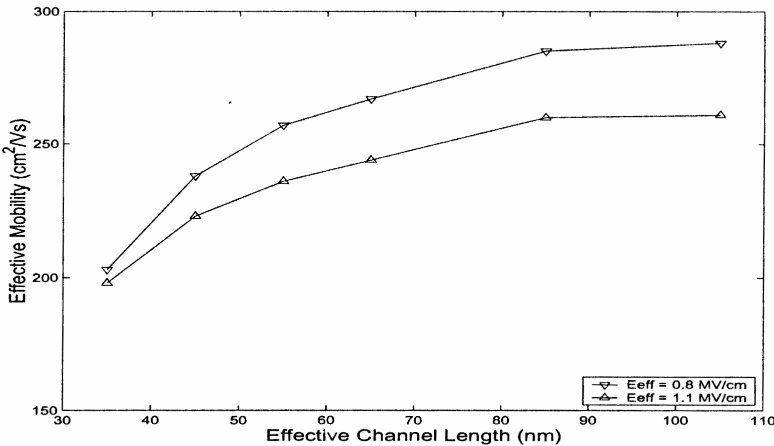

Figure 4. Measured μ_{eff} vs. L_{eff} for different E_{eff} on the t_{ox} = 1.5 nm technology.

As discussed earlier, the transmission coefficient, T, depends on the rate of momentum scattering and the width of the conduction band barrier at the source-end of the channel. It should therefore be closely related to the effective low longitudinal field mobility in that region. There are two reasons why the effective mobility would be decreasing with scaling: 1) In modern MOSFETs strong doping halos are used to improve electrostatic integrity and therefore coulomb scattering can increase to where it dominates the transverse-field-controlled surface scattering. 2) Remote scattering by source dopants as been pointed out as possible mechanism [14]. However, it is well known that all the standard methods for mobility characterization fail for modern

short-channel MOSFETs, so these effects have not been verified to date. In this work we take advantage of inverse modeling, and therefore knowledge of parasitic resistance and channel length to do such a characterization as a function of channel length in the t_{ox}=1.5 nm technology. Figure 4 shows the results of these measurements for two different effective transverse fields. The device details are given in the caption. It is quite clear that the effective channel mobility indeed decreases with channel length and in fact becomes independent of transverse field for the shortest devices. This suggests, that the surface scattering mechanism is no longer dominant for short lengths and indeed local and/or remote coulomb scattering dominates. It can be concluded then that conventionally extracted long-channel mobility is not sufficient for understanding transport in modern MOSFETs. The corollary to this observation is that in order to approach the ballistic limit it will be necessary to find methods to increase mobility in ultra-short-channel MOSFETs, either through the introduction of silicon-compatible high-mobility materials such as strained Si on relaxed SiGe [e.g. 15], or by the removal of dopants [e.g. undoped-channel double-gate structures], or a combination of both.

5 Conclusion

Inverse modeling of modern MOSFETs can be used not only as an engineering tool but also in order to shed light on transport mechanisms via simulation of devices with high fidelity to the real experimental ones. It is found that the energy balance model approximation is adequate for modeling a very broad range of devices with a unique set of model parameters. It is also found that source-channel injection velocity is still far below the thermal velocity limit even in sub-50-nm nMOSFETs, and the reason appears to be increased scattering from dopants near the source.

Acknowledgement
This work has been supported by DARPA

References

[1] Z. Lee et al., IEDM proceedings, p. 683, 1997.
[2] M. J. Van Dort et al., Solid-State Electronics, Vol. 37 No. 3, 1994, p 411.
[3] S. A. Mujtaba et al., NUPAD V, Honolulu, Hawaii, June 5-6, 1994
[4] A. Forghieri et al., IEEE Trans. CAD, Vol. 7 No. 2, Feb. 1988.
[5] S. M. Sze, Physics of Semiconductor Devices, 2nd ed.(New York:John Wiley & Sons, 1981).
[6] M. V. Fischetti, IEEE TED, Vol. 38, 1991, p 634.
[7] G. Shahidi, D. Antoniadis, H. Smith, IEEE EDL, Feb. 1988, p 94-96.
[8] F. Assad, Z. Ren, S. Datta, M. Lundstrom, P. Bendix, 1999 IEDM Digest, p 547-550.
[9] T. Mizuno, R. Ohba, Journal of Applied Physics Vol. 82 No. 10, Nov. 1997, p 5235-5240.
[10] A. Lochtefeld and D. A. Antoniadis, IEEE EDL, Vol. 22 No. 2, Feb. 2001, p 95.
[11] S. Y. Chou and D. A. Antoniadis, IEEE TED-34, 1987, p 448.
[12] H. Hu, J. Jacobs, L. Su, D. Antoniadis, IEEE TED, April 1995, p 669-677.
[13] Y. Taur, C. Wann, D. Frank, 1998 IEDM Technical Digest, p 789-792.
[14] M. V. Fischetti and S. E. Laux, J. of Applied Physics, Vol. 89 No. 2, Jan. 2001, p 1205.
[15] K. Rim, J. Hoyt, J. Gibbons, IEEE TED, July 2000, p 1406-1415.