On the Relationship Between Carrier Mobility and Velocity in sub-50 nm MOSFETs via Calibrated Monte Carlo Simulation

by

Osama Munir Nayfeh

B.S University of Illinois at Urbana-Champaign (2002)

Submitted to the

Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 2004
[June 2004]

Signature of Author _____
Department of Electrical Engineering and Computer Science, May, 2004

Certified by _____
Dimtri A. Antoniadis
Professor of Electrical Engineering
Thesis Supervisor

Accepted by _____
Arthur Smith
Professor of Electrical Engineering
Chair, Departmental Committee on Graduate Students

On the Relationship Between Carrier Mobility and Velocity in sub-50 nm NMOSFETs via Calibrated Monte Carlo Simulation

by

Osama Munir Nayfeh

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements
for the degree of
MASTER OF SCIENCE in
Electrical Engineering

Abstract

Subsequent to accurate 2D inverse modeling in the regime sensitive to electrostatics of industrial sub-50 nm NMOSFETs, a 2D full-band Monte Carlo device simulator was calibrated in the regime sensitive to transport parameters. The relationship between electron mobility and high-electric-field velocity at the source-channel potential energy barrier was investigated. The results show a strong correlation, as was demonstrated previously experimentally. Moreover, further proof is provided that the velocity at which carriers are injected from the source region in modern NMOSFET's is only about half of the limiting thermal velocity.

Thesis Supervisor: Dimitri A. Antoniadis

Title: Professor of Electrical Engineering

Table of Contents

List of Figures

Chapter 1

Introduction

The expression for drive current in a modern MOSFET, I_{on}, can be written as

$$I_{on} = WQ_i(x_0)\langle v(x_0)\rangle . \qquad (1)$$

Electrostatics determine Q_i (x_0), the density of carriers at x_0, the top of the source-to-channel potential energy barrier (conduction band peak), and transport parameters determine $\langle v(x_0)\rangle$, the average velocity of carriers at x_0, in the direction of the channel [1], the limit of which is the thermal velocity v_T where transport is purely diffusive. The thermal velocity is a function of the inversion charge density in the semiconductor 2D electron gas (inversion layer) [2]. The thermal velocity exceeds the saturation velocity v_{sat} (e.g. 1.0×10^7 cm/s for electrons) and it has been demonstrated previously experimentally that the carrier velocity at room temperature in silicon MOSFET near x_0 can exceed v_{sat}. This effect is called velocity overshoot [3].

The lateral electric field in a long-channel MOSFET is low, thus improving low-field mobility, μ, directly enhances $\langle v(x_0)\rangle$, hence I_{on} [4]. Low-field mobility is directly related to the inverse scattering rate τ by the expression $\mu=q/(m^*\langle\tau\rangle)$, where q is the elementary electron charge, m^* is the carrier conduction effective mass, and $\langle\tau\rangle$ is the average scattering rate. Under low-field conditions, carriers do not acquire substantial energy and $\langle\tau\rangle$ remains relatively constant. However, under high lateral-electric-field $>10^5$ V/cm in short-channel MOSFETs, carriers acquire enough energy to induce emission of optical phonons, which causes velocity saturation,

and also non-steady-state effects such as velocity overshoot which may occur due to the energy dependence of $<\tau>$ and the time it takes for steady state conditions to be achieved [5]. Despite the dependence of I_{on} to velocity saturation and overshoot in the high-lateral field regime, previous experimental results of enhancing mobility by mechanical bending have shown that there still exists a significant correlation between low-field mobility and high-electric field velocity even in deeply scaled NMOSFET devices [6].

Moreover, how close a modern industrial sub-50 nm MOSFET operates to the thermal limit, v_T has also been an issue of debate. Recently, experimental studies have shown that the near-source electron velocity in a sub-50 nm 1V bulk NMOSFET is less than 40% of the limiting thermal velocity and that continued device scaling does not seem to cause this velocity to approach the thermal limit [7].

In this study, 2D inverse modeling was used to fully characterize a sub-50 nm industrial silicon NMOSFET in the regime that is sensitive to electrostatics as described in Chapter 2, and this 2D inverse modeled structure was then used to calibrate scattering rates used in a Monte Carlo device simulator to fully match weak and strong inversion I-V characteristics under low and high-lateral field transport conditions, as described in Chapter 3. Calibrated Monte Carlo simulation with accurate electrostatics and transport was then performed in Chapter 4 to examine the previous experimental findings on how close a modern NMOSFET operates to the thermal velocity limit [7], and in Chapter 5, on the relationship between low-electric-field mobility and high-electric field velocity [6].

Chapter 2

Full 2-D Characterization by Inverse Modeling

2.1. Motivation of Inverse Modeling

In order to calibrate transport-sensitive Monte-Carlo scattering rates to the measured data of the sub-50 nm MOSFET examined in this thesis, a full numerical electrostatic representation of the device is first necessary [8]. This electrostatic representation includes information about the gate stack, as well as the 2D dopant profile. Extracting the elaborate 2-D dopant profile of a modern MOSFET (see Fig. 1.1), is not easy, and this problem has received much attention. Direct techniques such as 2-D SIMS (secondary ion mass spectroscopy) are difficult, destructive, and inaccurate in reproducing a 2-D profile with fine features. On the other hand, inverse modeling is an indirect and nondestructive method to extract the 2D dopant profile [8 a,b].

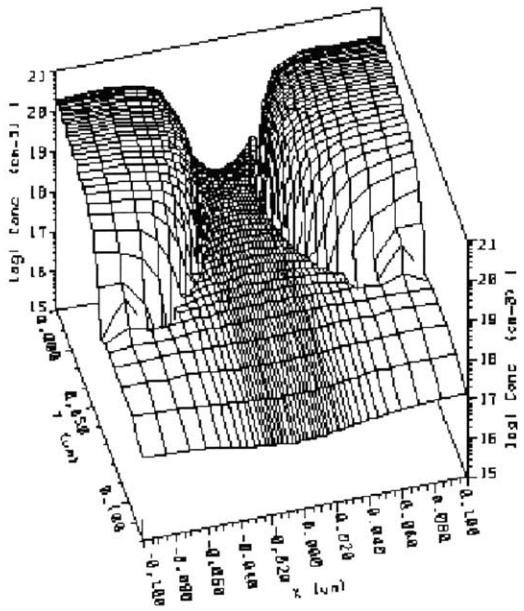

Fig. 2.1. Elaborate 2D doping profile in deeply scaled MOSFET [8].

Inverse modeling (IM) is a technique to fully characterize a device by fitting electrical measurements with simulation of a parameterized structure. Electrically measured device characteristics are matched to numerical simulation in the regime with low sensitivity to transport models and high sensitivity to electrostatics. IM requires a 2-D continuum device simulator, measured device electrical data (I-V, C-V, etc…), a parameterized 2-D structure, and a numerical optimizer. The detailed device parameters in the form of dimensions, concentrations, and accurate 2-D dopant profiles that are extracted by IM are seen in Fig. 1.2 below. [8].

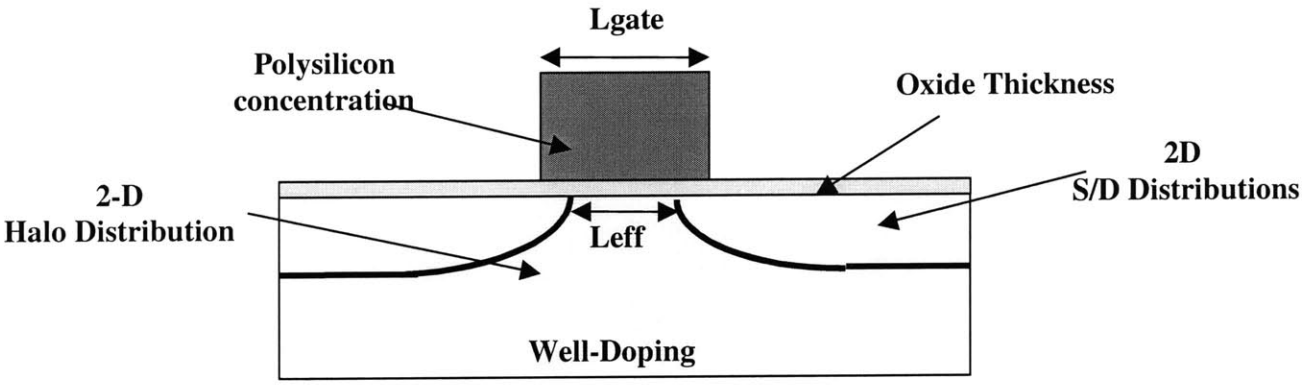

Fig. 2.2. Parameters extracted from IM

2D inverse modeling using a standard continuum device simulator was performed in the regime that is sensitive to electrostatics and insensitive to transport models, i.e. weak inversion, to fully characterize the industrial sub-50 nm "well tempered" NMOSFET studied in [6-7]. The procedure followed the flow-chart in Fig. 1.3. The device exhibits a subthreshold slope=95 mV/decade, and a drain induced barrier lowering, DIBL=100 mV/V.

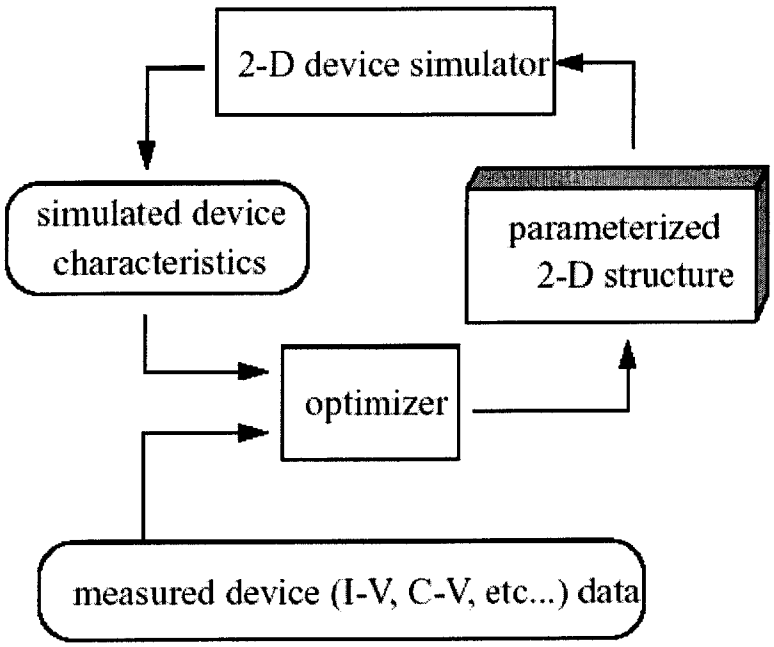

Fig. 2.3. IM flow-chart [8]

2.2 Gate-stack information

The first step in determining the full 2D numerical representation of a sub-50 nm MOSFET is to extract the gate-stack information. The physical gate oxide thickness and the gate polysilicon doping concentration are extracted by matching measured long-channel $C_{gg}=\partial Q_g/\partial V_g$ vs. V_{gs} under conditions of strong accumulation and strong inversion, with a parameterized numerical representation. A long-channel device of the same device family is used to minimize the effect of the relatively large overlap and fringing capacitances [8].

Under the regimes of strong accumulation and strong inversion, it is well known that the gate capacitance is mostly sensitive to the oxide thickness and the polysilicon doping concentration. Under strong accumulation conditions, $C_{gg\_acc.}=C_{ox}$, and the physical oxide thickness can be extracted by varying the physical oxide thickness to match C_{gg_acc} numerically. Under strong

inversion, $C_{gg\_inv}^{-1} = C_{ox}^{-1} + C_{poly}^{-1}$. C_{poly} is determined by the degree of polysilicon depletion, which is a function of the polysilicon doping concentration. Therefore, once the oxide thickness is extracted by matching C_{gg_acc}, the polysilicon doping concentration can then be extracted by fitting C_{gg_inv}, using the extracted value of T_{ox} [8].

C_{gg} vs. V_{gs} measurements were taken on a long channel NMOSFETs (to minimize relative fringing and overlap capacitances) of the same device family as the sub-50 nm NMOSFET examined in this thesis. The measurements were corrected for parasitic capacitances.

Due to the fact that in a sub-50 nm device, the gate oxide thickness is very thin (~1-2 nm), quantization effects in the MOSFET due to the wave nature of electron and holes cannot be ignored. The problem of quantization due to carrier confinement at the Si/SiO_2 interface and its effects on the charge distribution in a MOSFET inversion layer has been studied heavily over the last 2 decades [9]. The most basic quantization effects in MOSFETs are the shift of the threshold voltage and reduction of the gate capacitance.

Solution of the Schroedinger and Poisson equation has been shown to be an effective method to include these effects [9]. Unfortunately, solving these coupled equations can be extremely time consuming, so several approximate solutions have been proposed; the Van-Dort approximation and the density gradient method [10]. The density gradient method gives a more realistic representation of the carrier density that more closely mimics a Schroedinger solution.

The density gradient model was used to model quantization effects due to carrier confinement at the Si/SiO_2 interface in the IM of this thesis. The electron distribution determined by the density gradient model matches closely the Schrödinger solution.

The physical gate oxide thickness, T_{ox}=1.5 nm and the polysilicon doping concentration, $N_{poly}=6.25 \times 10^{19}$ cm^{-3} were determined as described above by matching the measured and

simulated $C_{gg}=\partial Q_g/\partial V_g$ under conditions of accumulation and inversion. The fit is shown in Fig. 2.3.

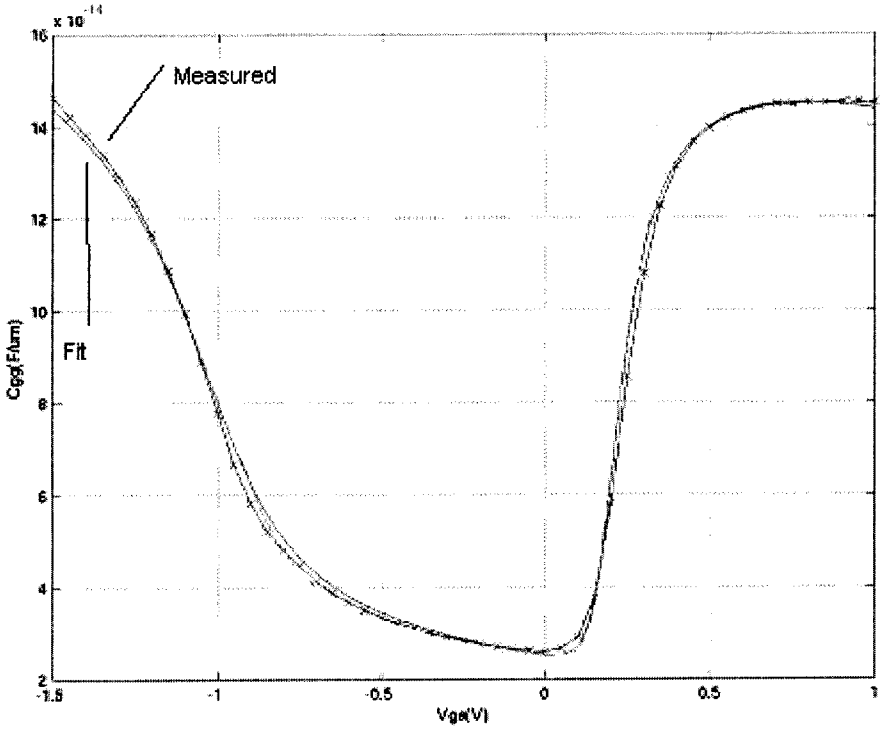

Fig. 2.4. Measured and Fit Cgg Vs. Vgs

2.3 Well-doping profile

The next step is to extract the well-doping profile. It has been shown that the log(Id) vs. Vgs characteristics under conditions of weak inversion are highly sensitive to the doping concentration and only weakly sensitive to transport models.

This sensitivity can be understood from:

$$I_D \approx WD\frac{dQ_i}{dx} = \frac{W}{L} I_x \exp\left(\frac{V_{GS} - V_x}{nV_{th}}\right)$$

$$I_x = \mu {V_{th}}^2 \frac{\gamma_{eff} C_{ox}}{2\sqrt{1.5\phi_F - V_{BS}}} \exp\left(-\frac{\phi_F}{2V_{th}}\right)$$

$$V_x = V_{FB} + 1.5\phi_F + \gamma_{eff}\sqrt{1.5\phi_F - V_{BS}} - \sigma_{DIBL}(N, V_{BS})V_{DS}$$

$$n = 1 + \frac{\gamma_{eff}}{2\sqrt{1.5\phi_F - V_{BS}}} \qquad \gamma_{eff} = A_{SCE}\frac{\sqrt{2\varepsilon_s qN}}{C_{ox}}$$

I-V subthreshold has strong dependence on 2-D doping through V_x, and a weak dependence on the mobility model [8].

In modeling the well-doping profile 3 gaussians are fit, and the standard deviation, peak location, and peak concentration are varied until a good match is made between measured and simulated log(Id) vs. Vgs, varying Vbs.

A retrograde well-doping profile was obtained by matching measured and simulated long-channel drain current (I_{ds}) versus gate-to-source voltage (V_{gs}) characteristics, by varying the back bias (V_{bs}) under conditions of weak inversion [8]. The fit is shown in Fig. 2.4, and the well-profile is shown in Fig. 2.5.

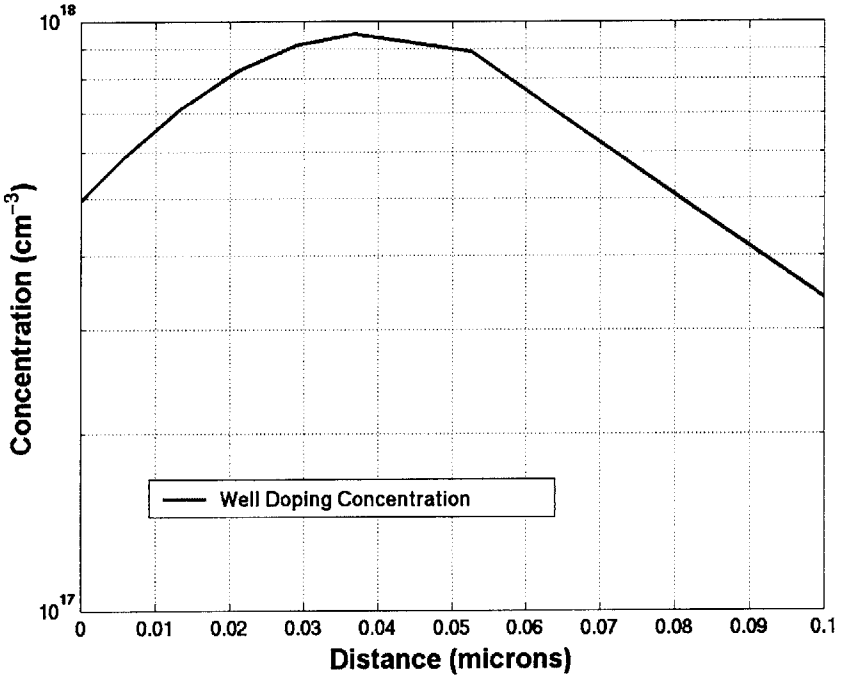

Fig. 2.6. Well doping profile extracted from fit in Fig. 2.5.

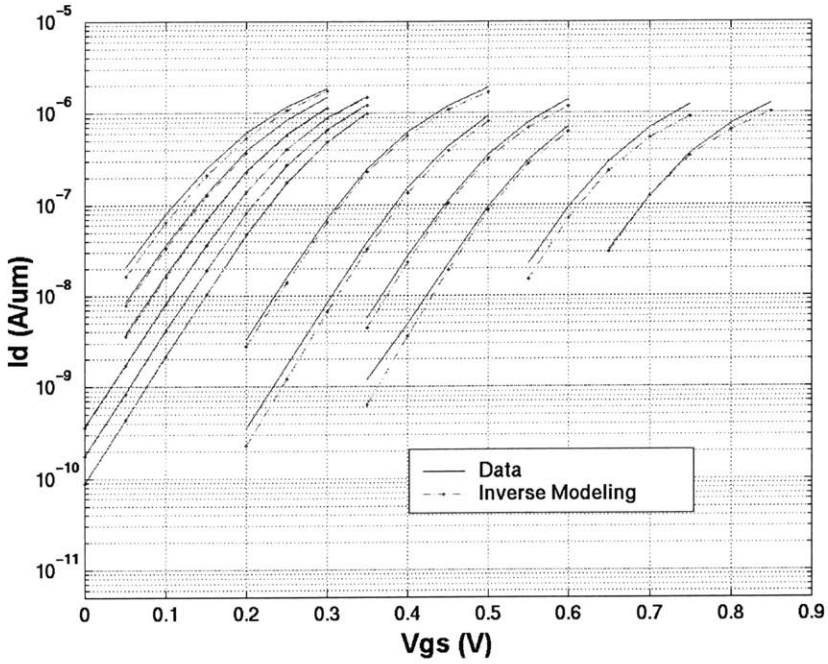

Fig. 2.5. Optimized subthreshold Id. Vs. Vgs. (measured and simulated), to extract well doping profile.

2.4 Halo and source/drain dopant profiles

Full short channel 2-D dopant profiles, including halo and source/drain, of a deeply scaled NMOSFET from the same device family were next determined by the coupled IV and CV inverse modeling method [11].

It was shown that short-channel C_{gsd} in accumulation is highly sensitive to the source/drain distribution, and short channel subthreshold log(Id) vs. Vgs is sensitive to the halo profiles, as it is a function of the amount of short channel effect [11].

Peak halo and source/drain doping concentrations of $4.7x10^{18}$ cm^{-3} and $2x10^{20}$ cm^{-3} respectively were determined. The effective channel length was then defined as the distance between source/drain junctions where the dopant concentration drops to $2x10^{19}$ cm^{-3}, and is ~40 nm for this deeply scaled device. The fits are shown in Fig. 2.6 (CV) and Fig. 2.7 (IV). Also a cut of the doping profile is shown in Fig. 2.8.

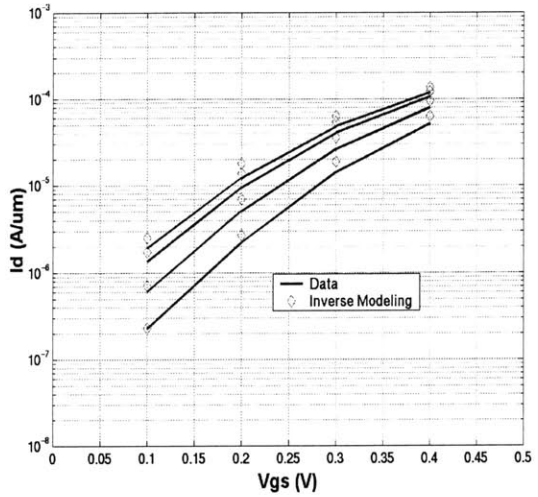

Fig. 2.7. Subthreshold I_d. Vs. V_{gs} (varying V_{ds}), short channel fit for extracting halo /SD doping profile.

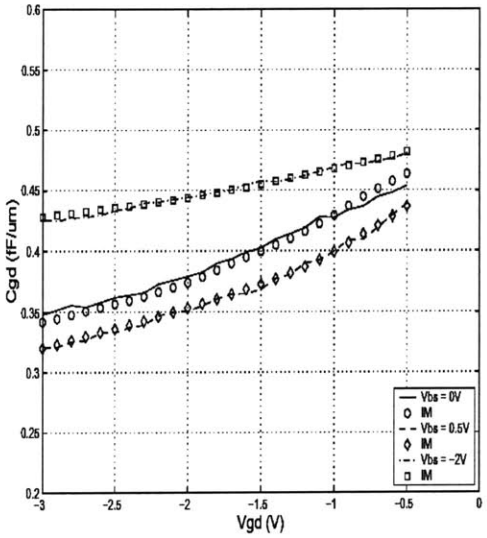

Fig. 2.8. C_{gsd} vs. V_{gs}. For short channel device, for determining details of SD/

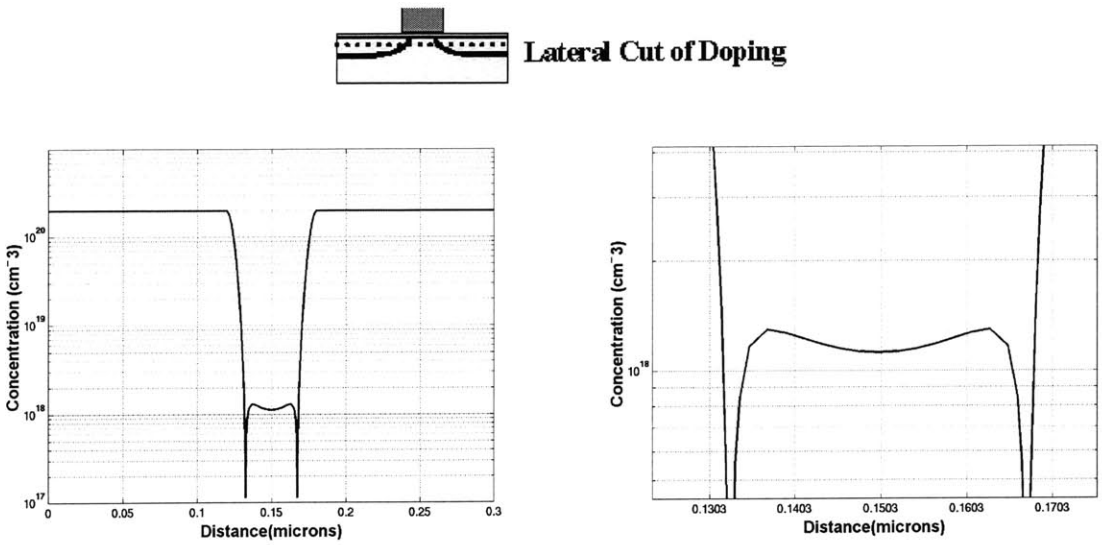

Fig. 2.9. Lateral cut of halo/SD doping and a zoomed in version on the right.

2.5 Extraction of Series resistance

A total source/drain series resistance R_{sd}=220 Ω·μm was subsequently determined by fitting measured data in strong inversion to simulation using a continuum device simulator that was calibrated for mobility. The measured I_{ds} vs V_{gs} data was then corrected for R_{sd} by defining a new $V_{gs}^{*} = V_{gs} - I_{ds} * R_{sd}$. A full contour plot of the 2D dopant profile, as well as other details of the NMOSFET features are shown in Fig. 2.10 [8].

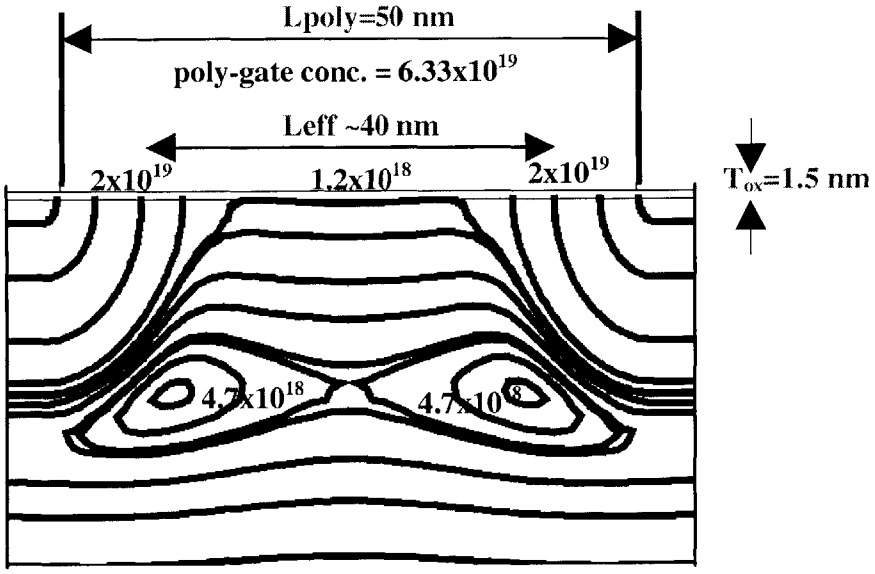

Fig. 2.10. Full 2D features of Leff ~40 nm device as determined from IM.

Chapter 3

CALIBRATION OF MONTE-CARLO DEVICE SIMULATOR:

3.1 Full Band Monte Carlo for semiconductor device simulation

In chapter 2, the technique of 2D inverse modeling was utilized to fully match the weak-inversion, Id vs. Vgs, and the capacitance vs. voltage characteristics of a sub-50 nm NMOSFET. These characteristics were shown to be highly sensitive to electrostatics and only weakly dependent on transport parameters. In order to fully match the full transfer characteristics of the device, in strong inversion and under applied lateral fields, transport parameters must be properly calibrated subsequent to the electrostatic modeling.

In this chapter, the semi classical technique of full-band Monte Carlo for semiconductor device simulation is first described. Next the different scattering processes included in the Monte Carlo device simulation are discussed briefly and calibrated when necessary to fully match the low and high lateral field Id. Vs. Vgs characteristics of the Leff~40 nm NMOSFET.

A solution to the semi classical transport problem of the Leff ~40 nm MOSFET is to solve the Boltzmann Transport Equation (BTE), coupled with a self-consistent solution of the Poisson equation. The BTE is an extremely difficult equation to solve directly. The 2 main approximations commonly implemented in a continuum numerical simulator are the Drift-Diffusion, and Hydrodynamic models. These approximations involve the solution of coupled partial differential equations, which can be solved numerically in 1, 2, and 3 dimensions.

However, these 2 approximations require the calibration of deterministic transport values such as the energy relaxation rate, mobility, and saturation velocity [12]. Also, the validity of these models begins to come into question for very shot channel MOSFETs. Subject to the extremely high lateral fields in a sub-50 nm MOSFET, the drift diffusion model is unable to capture nonlocal effects such as velocity overshoot. The hydrodynamic model with properly calibrated values continues to be valid down to 20 nm channel length, but is questionable in device that operate near the ballistic limit. Moreover, due to the difficulty of calibration of these high-field transport parameters and to the desire to obtain mode physical information at all regions of operation, the technique known as Monte Carlo device simulation is used in this thesis [8].

The combination of a statistical solution of the Boltzmann Transport Equation and a self-consistent solution of the Poisson Equation continues to gain popularity as a viable approach to device simulation. Monte Carlo simulation provides a statistical solution to the semi classical transport problem by tracking individual electrons as they move in space and time in the crystal lattice [12].

In the Monte Carlo technique, the physics enters though the description of the semiconductor band structure (i.e. kinematics) and the scattering behaviors (i.e. dynamics). In the deterministic approaches, the physics is lumped into parameterized motilities, diffusion constants, and lifetimes. Full-band Monte Carlo simulation provides detailed information on the distribution function, which can only be obtained by solving the Boltzmann transport equation (BTE) statistically by Monte Carlo techniques. This distribution function then can provide accurate energy-dependent scattering rates at all positions in a MOSFET. However, in exchange for this advantage, significantly longer computational execution times are required [13].

The Monte Carlo device simulator used in the simulations in this thesis, is MOCA, which was developed at the University of Illinois at Urbana-Champaign. MOCA uses the full semiconductor band structure because the lateral-electric-field in a sub-50 nm MOSFET in saturation is very high, $>10^5$ V/cm and is also rapidly varying over short distances. Under these conditions, carriers traveling in a MOSFET channel from source to drain, may gain substantial amounts of energy that causes population of **k** states not just in the vicinity of the first conduction band minimum. Since scattering rates are dependent on the density of states and effective masses, both of which are determined by semiconductor band-structure, details of the silicon full-band structure must be taken into account in order to accurately model I_{on} via Monte Carlo simulation [14].

Fortunately, advances in solid state physics in the 1950's and 1960's have provided techniques of determining the accurate full semiconductor band structure. The main methods of determing full band structure are the tight binding, and the pseudopotential. The pseudopotential technique is used to determine the full-band structure used in MOCA. Fig. 3.1 shows the Silicon band structure at important symmetry points [12].

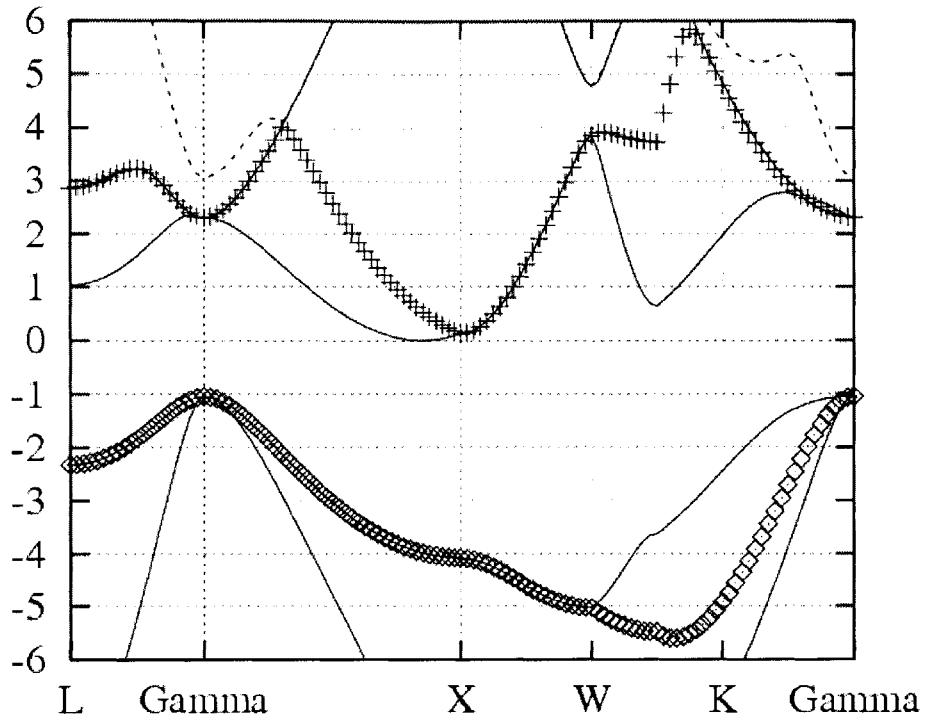

Fig. 3.1. Silicon band structure showing important symmetry points, as calculated from the pseudopotential technique. More information on the details can be found in [12]

3.2 Calibration of Scattering Rates to Measured Data using Inverse Modeled Device Structure

The full-band Monte Carlo simulator (MOCA) considers f and g type intervalley acoustic phonon scattering, X-X intervalley phonon scattering, X-L intervalley phonon scattering, surface roughness scattering, and ionized impurity scattering [14]. Quantum mechanical effects are included via self-consistent coupling of the Schrödinger equation [15]. In the following Monte Carlo calibration, only adjustments to the surface scattering rates were needed to fully fit by Monte Carlo simulation measured low and high lateral field I_{ds} vs V_{gs} characteristics of the sub-50 nm NMOSFET using the device structure from IM shown in Fig. 2.10.

Every device family has a unique Si/SiO_2 interface, thus the surface roughness scattering rate is also unique. The surface roughness scattering rate is derived from a statistical exponential auto covariance function [16]

$$\delta(r) = \Delta^2 e^{-r/L} \qquad (2)$$

Δ is the rms height of the amplitude roughness and L is the correlation length. Fourier transformation of (2) then yields the power spectrum density [16]

$$S_{sr}(q) = \pi\Delta^2 L^2 \cdot \frac{1}{[1+(L^2 q^2/2)]} \qquad (3)$$

where $q^2 = 2k^2(1-\cos\theta)$, where θ is the scattering angle (away from surface plane). The surface scattering rate, τ_{sr} can then be derived from the power spectrum density (3) and is:

$$\frac{1}{\tau_{sr}(\mathbf{k}_{//})} = \frac{m_d e^2 E^2_{eff} \Delta^2 L^2}{2\hbar^3} \times \int_0^{2\pi} \frac{1}{1+L^2 q^2/2}(1-\cos\theta)d\theta \qquad (4)$$

where $\mathbf{k}_{//}$ is the wave vector, $\hbar$ is the reduced Planck constant $E_{eff} = \frac{e}{\kappa_{si}\varepsilon_0}(\frac{1}{2}N_s + N_{depl})$, is the effective vertical field in the channel, where N_s is the inversion charge density, and N_{depl} is the depletion charge density [16].

The statistical parameters Δ and L of (2) were adjusted to calibrate the surface roughness scattering rates for this particular Si/SiO2 interface, so that a good match is made between the

measured and Monte Carlo simulated low-lateral field I_{ds} vs. V_{gs} characteristic of the Leff ~40 nm NMOSFET under conditions of strong inversion, sensitive to surface scattering. The measured device characteristics have been adjusted for the source/drain series resistances R_{sd} as described earlier. The fit is shown in Fig. 3.2 and shows good agreement in both weak inversion, which is sensitive to electrostatics and is set by 2D inverse modeling, and in full inversion, which is sensitive to low-field transport, and is set by mobility.

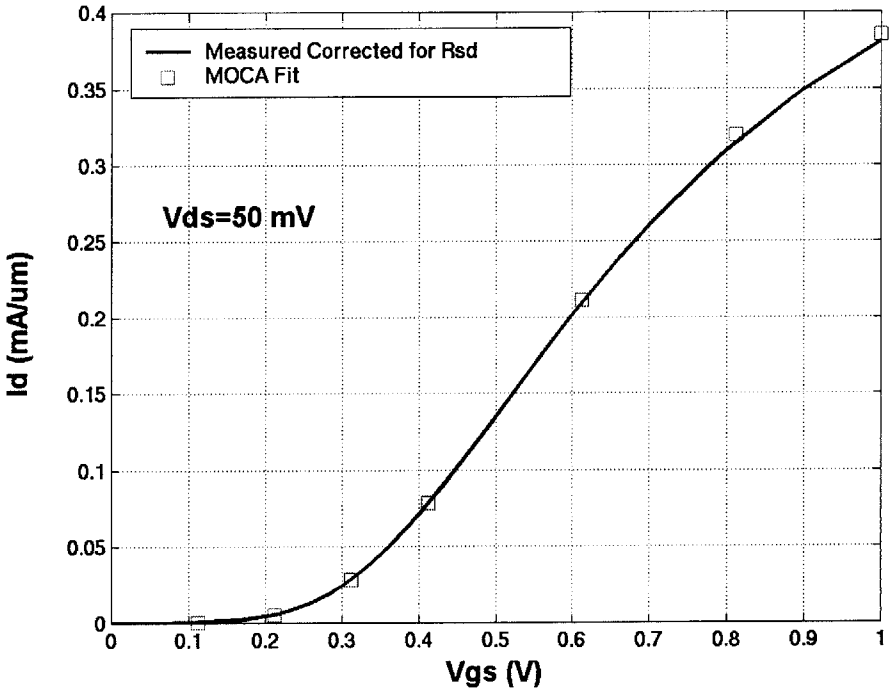

Fig. 3.2. Low-lateral field Id. Vs. Vgs (measured and fit). The measured data is corrected for Rsd as described in Chapter 2. Good agreement is shown.

Screened impurity scattering rates in MOCA are calculated using the three-body exclusion Ridley statistical model [17]. No modifications were needed to the impurity scattering model in the fit shown in Fig. 3.3. Moreover, the Ridley screened statistical impurity scattering model used was compared to a screened impurity analytical scattering model independently developed

from experimental data [18]. Fractional screened impurity scattering rates as a function of integrated channel charge N_i for a constant channel doping level of $2.5x10^{18}$ cm^{-3} were compared between the Ridley statistical model used in the Monte Carlo simulations in this paper, and the analytical model in [18] which is based on experimental data. The fractional screened impurity scattering rate is defined as τ_{imp}/τ_{tot} where τ_{imp} is the impurity scattering rate and τ_{tot} is the total scattering rate. The results are in good agreement as shown in Fig. 3.4

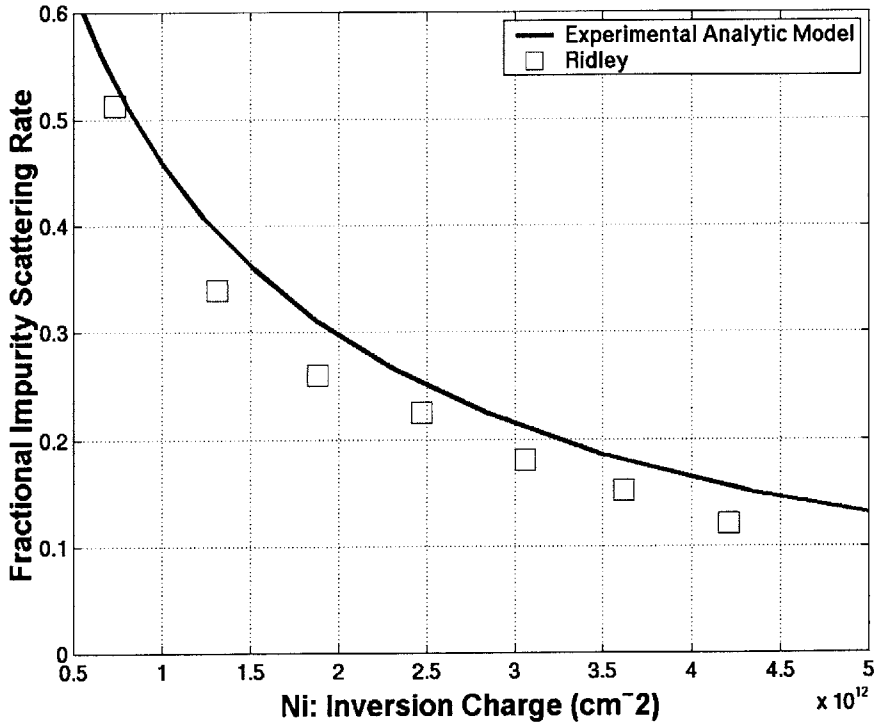

Fig. 3.3. Fractional screened impurity scattering rates versus inversion charge density for an experimentally developed screened impurity model [18] and in the model used in the Ridley statistical impurity scattering model.

The acoustic phonon parameters used to compute phonon scattering rates are those referenced by Tang and Hess [19], and are listed in Table 1. No changes were needed to the acoustic phonon parameters in the low-field fit in Fig. 3.2.

In order to accurately model I_{on} in the high-lateral field regime, where carrier saturation and velocity overshoot occur, accurate optical phonon parameters must be used [19]. Using the calibrated scattering rates from the low-lateral field fit (surface, impurity, and acoustic phonon), and the default optical phonon parameters from Tang and Hess, shown in Table 1, a good match is obtained between the measured I_{ds} vs V_{gs} characteristics corrected for R_{sd} under high lateral field conditions (V_{ds}=1.0 V) and Monte Carlo simulation. The fit is shown in Fig. 3.3.

Table 1 Phonon Parameters from Tang & Hess [ ]		
Phonon Temp. (K)	**Deformation Potential (10^7 eV/cm)**	**Scattering Type**
X-X Intervalley scattering		
220	3	f
550	20	f
685	20	f
140	5	g
215	8	g
720	110	g
X-L Intervalley scattering		
672	25	
634	25	
480	25	
197	25	

Table 1. Acoustic and optical phonon parameters from Tang and Hess[19]

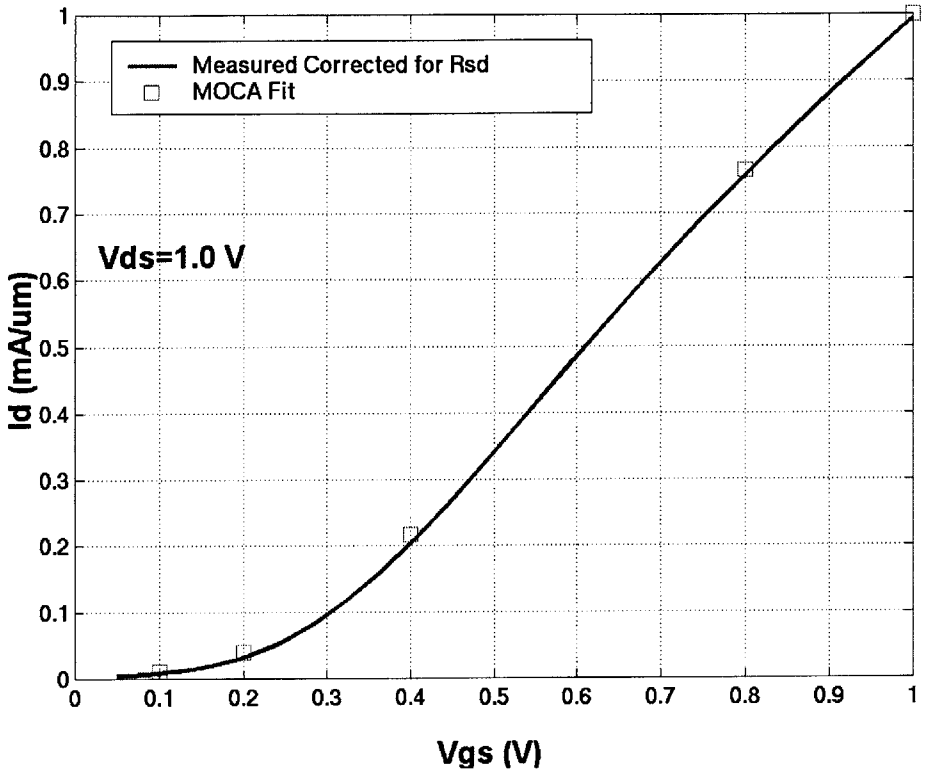

Fig. 3.4. High-lateral field Id. Vs. Vgs (measured and fit). The measured data is corrected for Rsd as described in Chapter 2. Good agreement is shown.

Thus, only calibration of the surface roughness scattering rates was required to match Monte Carlo simulation of the inverse modeled device structure in Fig.2.10 with measured I_{ds} vs V_{gs} characteristics under low and high-lateral field transport conditions.

Chapter 4

HOW CLOSE TO THE THERMAL LIMIT

4.1 Previous Experimental Results

One way of improving the performance of modern MOSFETs, is to increase the source injection velocity. However, as described earlier in Chapter 1, there are fundamental limits to the maximum achievable value of the source injection velocity, the thermal velocity, which is set by semiconductor physics. Thus if a modern MOSFET already operates close to this limit, continued scaling should not provide substantial increases in velocity, hence drive current. This concept is especially important in the design of new generation MOSFETs, especially when considering using mobility-enhanced materials in hope of increasing carrier velocity. Therefore, it is extremely important to determine how close a modern MOSFET operates to the thermal velocity limit.

A direct experimental method of determining the source injection velocity close to the conduction band peak was demonstrated in [7]. It was shown for the sub-50 nm industrial MOSFET modeled in this thesis that the source injection velocity is far from the thermal velocity limit, and thus continued scaling should not bring this velocity closer to this limit (See Fig. 4.1).

Due to many uncertainties, directly determining the carrier velocity close to the conduction band peak of a sub-50 nm MOSFET is a difficult task. Large relative fringing and overlap capacitances, and nonuniform inversion channel density complicate measurements. Nevertheless, a reliable experimental method of measuring the electron velocity very close to the conduction

band peak was demonstrated previously. In the experimental method, $v_{eff}=I_{on}/(WQ_i(x_0))$, is determined directly from a deeply scaled device [7].

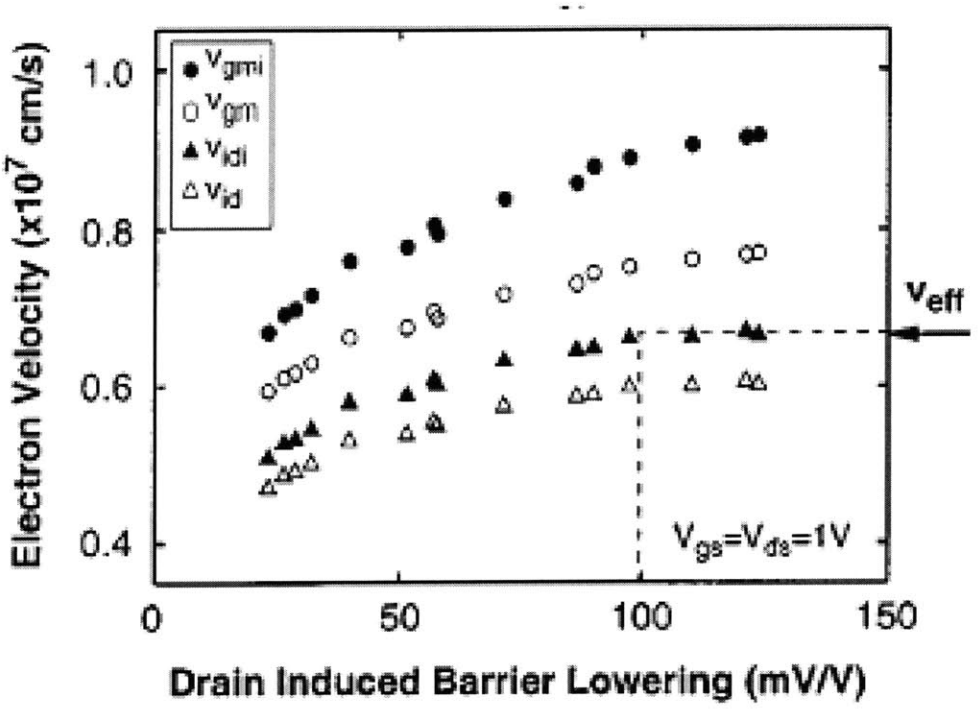

Fig. 4.1. Previous experimental results on extracting velocity in deeply scaled MOSFETs

In this chapter Monte Carlo simulation is performed to confirm the experimental results, namely that this velocity is only ~ 50% of the limiting thermal velocity.

4.2 Mean Free Path Determination from Accurate Scattering Rates

The degree of backscattering of carriers to the source determines how close the actual velocity at the conduction band peak of a deeply scaled NMOSFET is to the limiting thermal velocity [5]. In the absence of backscattering, $<v(x_0)>$ approaches v_T and transport is "ballistic". In the regime

of ballistic transport, the equation for Ion (1) can be written as $I_{on}=I_{ballistic}=W\cdot Q_i(x_0)\cdot v_T$. Ballistic transport is expected to occur when the mean free path for scattering in the conduction band peak region is less than the critical distance of a k_BT/q potential drop [5].

To determine reliably the position dependent vector mean free path, accurate scattering rates across the length of the device are needed. As seen in Fig. 4.2, which shows the scattering rates determined by calibrated Monte Carlo simulation along the length of the NMOSFET structure of Fig. 2.10, many scattering events may occur along the length of the device. The scattering rates τ, were computed as a position dependent value $\tau(y)$, weighted by the inversion charge density $n(y)\cdot$ and is given by $\tau=(\tau(y)\cdot n(y)\cdot dy)/(n(y)\cdot dy)$. The position dependent vector mean free path in the direction of the channel λ is calculated across the length of the device as $\lambda=<(1/\tau)\cdot v>$, and is shown in Fig. 4.3. This value <1 nm in the conduction band peak region is much less than the critical length of the k_BT/q potential drop for this device, ~5 nm. We thus can conclude that this real industrial "well tempered" device operates far from the ballistic limit.

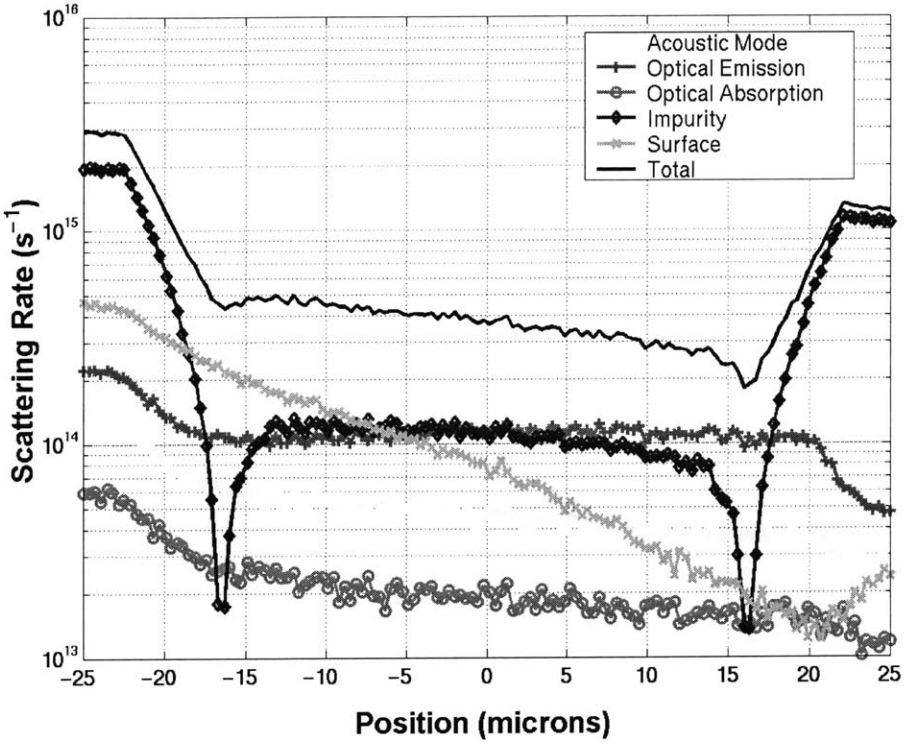

Fig. 4.2. Scattering rates across the length of the inverse modeled device in Fig. 1 under high-lateral field conditions and full inversion.

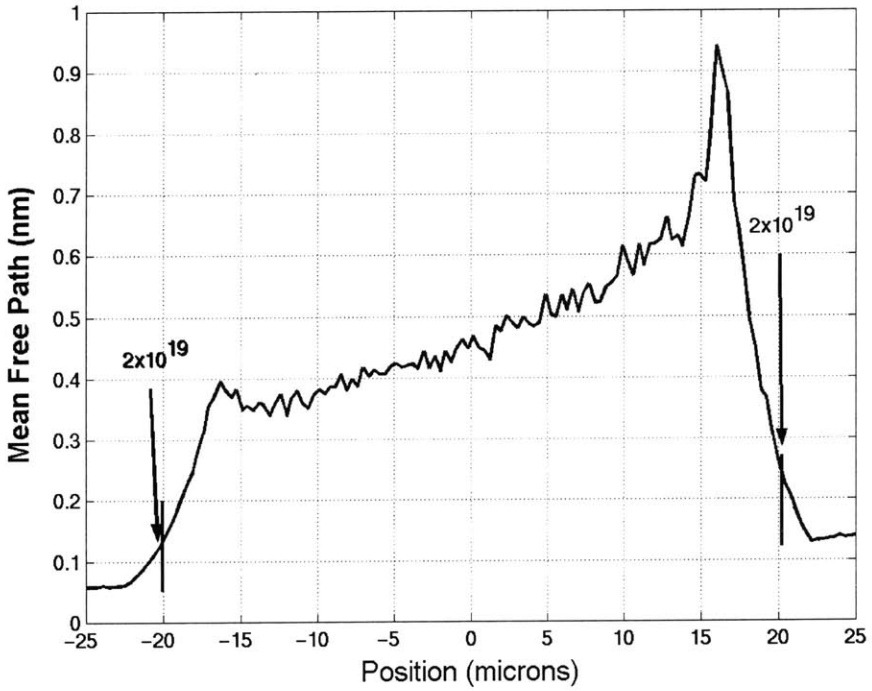

Fig. 4.3. The vector mean free path λ across the length of the device. Also shown are the points in the device where the doping=$2x10^{19}$ cm^{-3}.

As seen in Fig. 4.2, a large fraction of the total scattering rate is due to impurity scattering and surface scattering. The high fraction of impurity scattering is due to the large halo doping values needed to suppress short channel effects in order to maintain the "well tempered" behavior, and the high surface scattering rate is due to the very high effective vertical field in the inversion layer. Continued scaling of a bulk MOSFET increases the short channel effects further, and thus an even higher impurity concentration will be required, thus increasing the impurity scattering rate further.

4.3 Comparison with thermal velocity limits

Calibrated Monte Carlo simulation was performed in the high-lateral field regime with nominal ($V_{gs}=V_{ds}=V_{DD}=1V$) and the average velocity of carriers at the source-channel potential energy barrier, $\langle v(x_0)\rangle$, was calculated as a weighted average of the position dependent velocity and inversion charge density $\langle v(x_0)\rangle = (v(y)\cdot n(y)\cdot dy)/(n(y)\cdot dy)$. Since the thermal velocity v_T is dependent on inversion charge density [5], the gate-source voltage V_{gs} was then varied while keeping $V_{ds}=1V$, and the value of $\langle v(x_0)\rangle$, for the Leff ~40 nm NMOSFET, determined by Monte Carlo simulation, was compared with the theoretical thermal velocity limit as a function of inversion charge density N_i [], as shown in Fig. 4.4. The values of the thermal velocity v_T were obtained by reducing the MOCA scattering rates in the channel until ballistic conditions were achieved. The degree of backscattering to the source can be determined as $\langle v(x_0)\rangle/v_T$ as shown in Fig. 4.5, and this value determines how close the device operates to the thermal limit v_T.

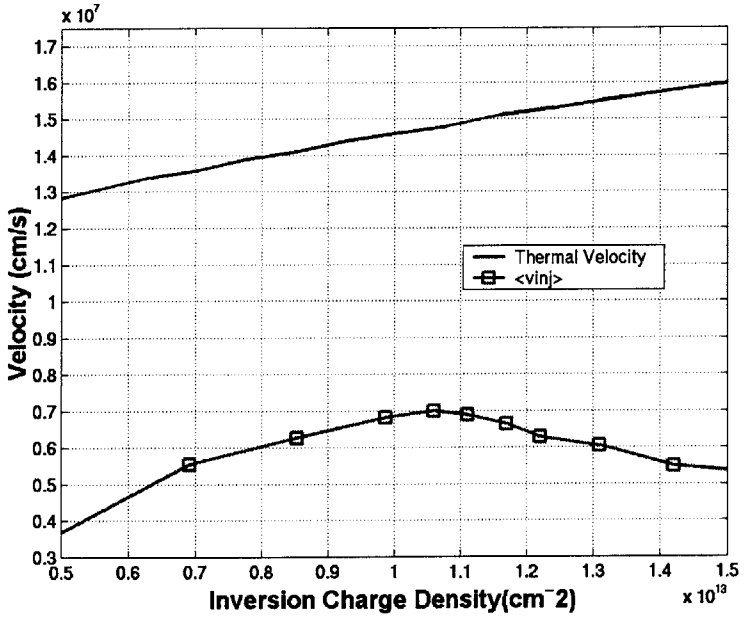

Fig. 4.4. Average velocity of carriers at the conduction band peak $\langle v(x_0)\rangle$ and the thermal velocity limit v_T as a function of inversion charge density.

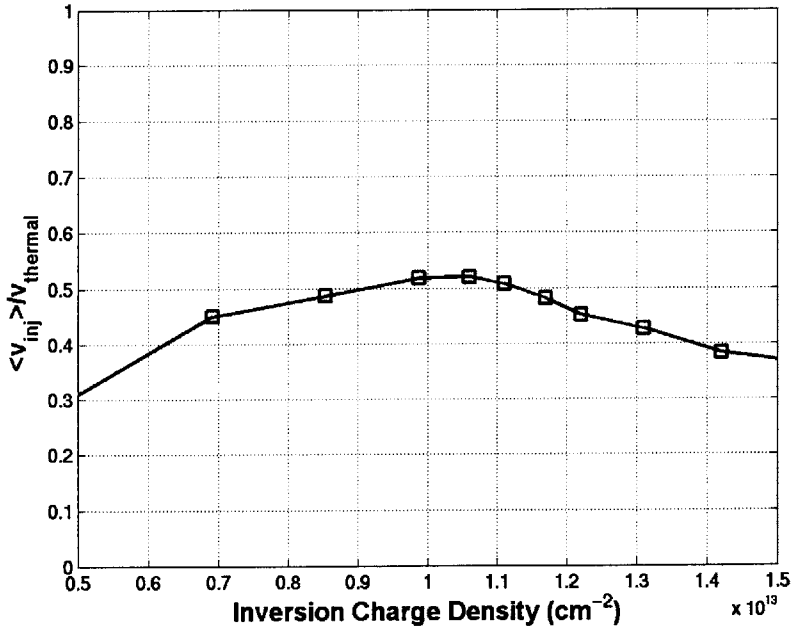

Fig. 4.5. The ratio $\langle v(x_0)\rangle/v_T$ as a function of inversion charge density. This value represents the amount of backscattering to the source, and is a measure of how close the device operates to the thermal limit.

Clearly it is seen from Fig. 8, that this deeply scaled device operates at only 45-55% of the thermal velocity limit. The somewhat larger values from this study compared to experiment are due to the fact that in the experimental work (40%) a single value for the thermal velocity was used ($1.7x10^7$ cm/s). Nevertheless, the general result that a modern NMOSFET only operates at around half of the thermal limit is verified.

Chapter 5

RELATIONSHIP BETWEEN LOW-FIELD CARRIER MOBILITYAND HIGH-FIELD VELOCITY

5.1 Previous Experimental Results

Previously, uniaxial strain applied via mechanical bending of deeply scaled NMOSFETs demonstrated a significant relationship between low-field mobility and high-electric field velocity. A schematic of the bending procedure shown in Fig. 5.1, and Fig. 5.2 shows the experimental results. [6].

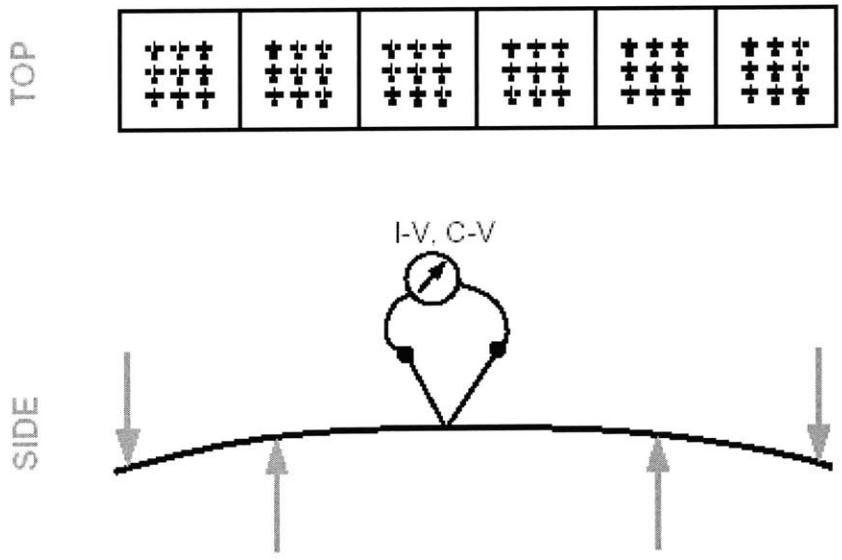

Fig. 5.1. Schematic of bending procedure used to investigate relationship between mobility and velocity in previous study.

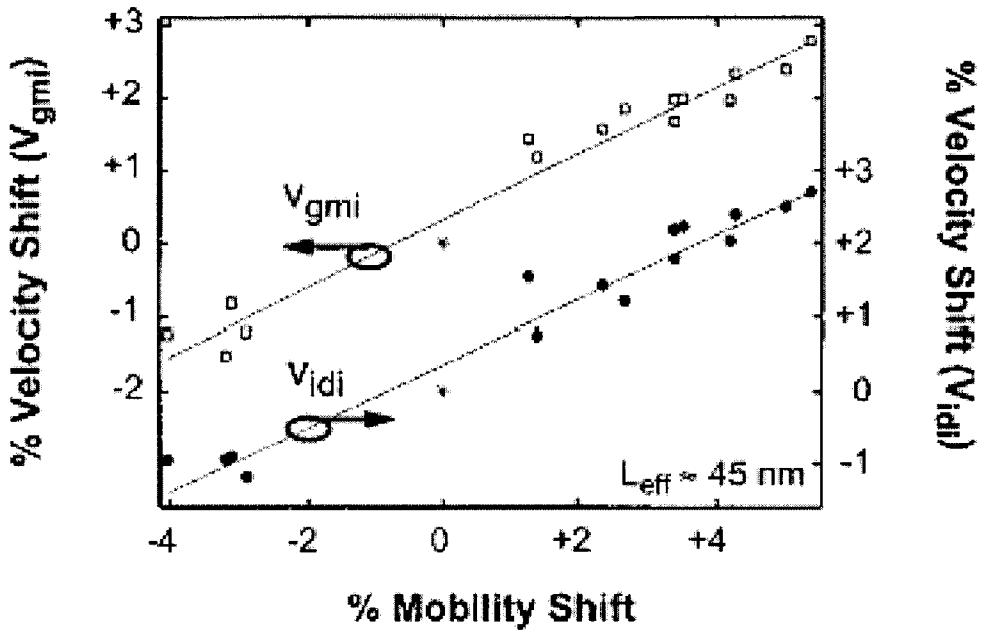

Fig. 5.2. Experimental results on relationship between low-field mobiity and velocity.

It was shown that the measured dependence of near- source electron velocity $\langle v(x_0)\rangle$ on low-field mobility μ_{eff}, defined by the ratio $R_{v,\mu}=\delta\langle v(x_0)\rangle/\delta\mu_{eff}$ = 0.44 -0.46 [6]. For example, $R_{v,\mu}$=0.45 implies that a silicon channel with a 10% low-field mobility increase will exhibit ~4.5% increase in I_{on}. We performed the full-band calibrated Monte Carlo simulation, as discussed earlier, to compare with the experimental relationship between low-field mobility and velocity in the saturation regime.

5.2 Enhancing Mobility by Altering the Scattering Rates

The surface-roughness scattering rate parameters Δ and L of eq. 3 were modified from their calibrated values to give a 13% change in the effective short-channel mobility (μ_{eff}) for the L_{eff}

~40 nm deeply scaled inverse modeled NMOSFET device shown in Fig. 1. No changes were made to the saturation velocity (i.e. optical phonon parameters), or to the impurity scattering. We justify our approach in comparing the effect of uniaxial strain via mechanical bending on a MOSFET, with a surface mobility increase, due to the small perturbation caused by the mechanical bending. A more proper approach would be to recalculate the band structure with strain. The impact of the surface mobility enhancement on the surface scattering rate and total scattering rate in the device is shown in Fig 5.3.

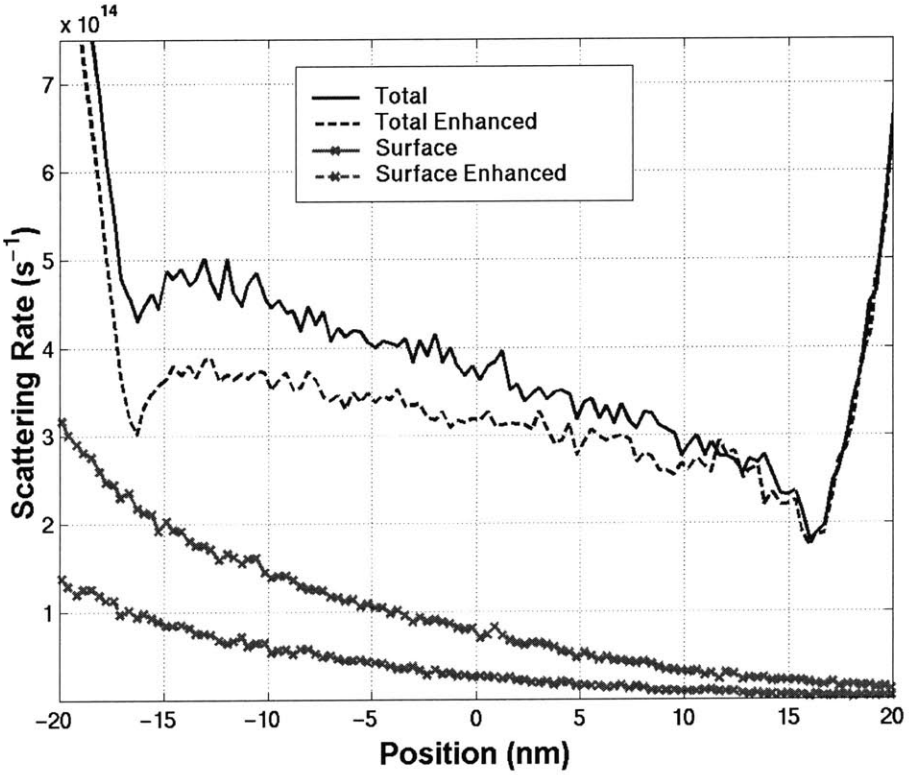

Fig. 5.3. Surface scattering and total scattering rate across the length of the inverse modeled device in Fig. 1 with enhanced surface mobility.

5.3 Comparison with previous experimental results

Despite the reduction in surface scattering, the mean free path is still much smaller than the the critical length of the k_BT/q potential drop, and the effective short-channel mobility change can be defined as the percent change in current in the linear regime $\delta\mu_{eff} = \Delta Id_{lin}/Id_{lin}$ at a gate-to-source voltage V_{gs}=1V, with a drain bias V_{ds}=50mV [8]. A drain-source bias V_{ds}=1.0 V was then applied to simulate the device in the high-electric field regime, and $\delta\langle v(x_0)\rangle = \Delta Id_{sat}/Id_{sat}$= was evaluated. Therefore Monte-Carlo simulation gives a value of $R_{v,\mu}=\delta\langle v(x_0)\rangle/\delta\mu_{eff}$=0.43, which agrees well with the value of $R_{v,\mu}$=0.44-0.46 from the previous experimental results.

Therefore, for the Leff ~40 nm device analyzed in this thesis, which only operates ~50 % to the thermal limit, low-field mobility enahcenment still plays an important role in increasing the drive current. However, this is not a one-to-one correlation, and the shortcoming can be explained by the fact that there is a high optical phonon emission scattering rate (not induced in low-lateral-field conditions), which induces carrier saturation, Moreover, the device operation is subject to the non-equilibrium transport phenomenon, velocity overshoot, which strength depends greatly on the energy relaxation time, and is not directly correlated to low-field mobility. Therefore, we should not suspect that devices that operate closer to the ballistic limit, to have such a strong correlation to low-field mobility as also reported in [5]. More deeply scaled devices will begin to be dominated more greatly by the optical emission processes (due to higher lateral fields), as well as the velocity overshoot effect.

Chapter 6

CONCLUSION

In this study, 2D inverse modeling was utilized to calibrate scattering rates used in a Monte Carlo device simulator. Moreover, we have used this calibrated Monte Carlo device simulator to confirm previous experimental results on two elusive issues, how close a modern industrial NMOSFET operates to the thermal limit, and on the relationship between low-field mobility and high-field velocity.

REFERENCES

[1] Mark Lunstrom and Zhibin Ren, "Essential physics of carrier transport in nanoscale MOSFET's," IEEE Transactions on Electron Devices, vol. 49, pp. 133-141, Jan 2002.

[2] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor, J. Appl. Phys., vol.. 76, pp. 4879-4890, 1994.

[3] G. Shahidi, D. Antoniadis, and H. Smith, "Electron velocity overshoot at room and liquid nitrogen temperatures in silicon inversion layers," IEEE Electron Device Lett., pp. 94–96, Feb. 1988.

[4] Mark Lundtrom, "Fundametals of Carrier Transport," Cambridge University Press, 2000.

[5] Lundstrom, Mark, "On the Mobility Versus Drain Current Relation or a Nanoscale MOSFET," IEEE Electron Device Lett., vol. 22, pp. 293-295, June 2001.

[6] A. Lochtefeld and D. Antoniadis, "Investigating the relationship between electron mobility and velocity in deeply scaled NMOS via mechanical stress," IEEE Electron Device Lett., vol. 22, pp. 591-593, Dec. 2001.

[7] A. Lochtefeld and D. Antoniadis, "On experimental determination of carrier velocity in deeply scaled NMOS: how close to the thermal limit,?," IEEE Electron Device Lett., vol. 22, pp. 95-97, Feb. 2001.

[8] Djomehri, Ihsan, "PhD thesis," Massachusetts Institute of Technology, 2002

[8b] Lee, Zach, "PhD thesis," Massachusetts Institute of Technology, 1999

[9] Nayfeh, O.M. "Influence of Polysilicon quantization effect on 2D advanced device characterization," SRC Review Presentation 2003

[10] ISE-8.0 Dessis, Manual

[11] I. Djomehri and D.A. Antoniadis, "Inverse modeling of sub-100 nm MOSFET's using I-V and C-V," IEEE Transactions on Electron Devices, vol. 49, pp. 568-575, April. 2002.

[12] K. Hess. Monte Carlo Device Simulation: Full Band and Beyond. Kluwer Academic Publishers, 1991.

[13] Christoph Jungemann, Bernd Meinerzhagen "Hierarchial Device Simulation," Springer Engineering 2003.

[14] Amanda Duncan, Umberto Ravaioli, and Jürgen Jakurmeit, "Full band Monte Carlo investigation of hot carrier trends in the scaling of metal-oxide-semiconductor field effect transistors," IEEE Transactions on Electron Devices, vol. 45, pp. 867-876, April. 1998.

[15] B. Winstead, H. Tsuchiya, U. Ravaioli, "An Approach to Quantum Correction in Monte Carlo Device Simulation," Nanotech 2001, pp. 566-569. Vol. 1.

[16] Yamakawa, et al., Miyatsuji, Masaki, and Ravaioli, "Study of interface roughness dependence of electron mobility in Si inversion layers using the Monte Carlo method," Journal of Applied Physics, vol. 79, pp. 911-916, Jan. 1996.

[17] B.K. Ridley, "Reconciliation of the Conwell-Weisskopf and Brooks-Herring formulae for charged-impurity scattering in semiconductors: Third-body interference," J. Phys. C, vol. 10, pp. 1589–1593, 1977.

[18] H.M..Nayfeh, C.W. Leitz, A.J. Pitera, E.A. Fitzgerald, J.L. Hoyt, D.A. Antoniadis, "Influence of high channel doping on the inversion layer electron mobility in strained-silicon n-MOSFETs," IEEE Electron Device Lett., vol 24, pp. 248-250, April 2003.

[19] J.Y. Tang and Karl Hess, "Impact ionization of electrons in silicon (steady state)," Journal of Applied Physics, vol. 54, pp. 5139-5144, Sep. 1983.