ABSTRACT

MORGENSEN, MICHAEL P. Drain Current Modeling and Characterization using MISFETs. (Under the direction of Dr. Leda Lunardi and Dr. Doug Barlage.)

The effects of interface states on metal-insulator-semiconductor field-effect transistor (MISFET) drain current have been investigated using the one-dimensional solution of Poisson’s equation commonly used to develop compact models. The U-like distribution of interface state density throughout the bandgap has been utilized when calculating interface state charge density without resorting to the numerous approximations found in the literature. Consequently, the resulting set of charge expressions are able to correctly predict gate voltage stretch-out in the current-voltage domain. A compact model of trap-affected drain current using the charge sheet approximation has been developed and the challenges highlighted. The computational speed needed by the model is a limiting factor and is caused by the unique relationship between inversion charge density and surface potential not found in standard models excluding interface states.

A new analytical correction to effective mobility measurements has been derived for the case of a bulk MISFET with negligible interface state density. The correction is capable of removing artificial low field roll-off from effective mobility results when the drain current is measured at a single drain-to-source voltage. Limitations imposed by high interface state density for effective mobility analysis have been discussed and a solution proposed. It has been found that a manipulation of drain current measured in the sub- and near-threshold regions can give insight into both the effective mobility and the parameters describing the interface state distribution. This approach is made possible by comparison with theoretical expressions governing the set of MISFET charges that include interface state charge density.
Drain Current Modeling and Characterization using MISFETs

by
Michael P. Morgensen

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Electrical Engineering
Raleigh, North Carolina
2011

APPROVED BY:

______________________________  ________________________________
Dr. Griff Bilbro                  Dr. Mark Johnson

______________________________  ________________________________
Dr. Doug Barlage
Co-chair of Advisory Committee   Dr. Leda Lunardi
Chair of Advisory Committee
BIOGRAPHY

Michael Morgensen was born in High Point, North Carolina to Thomas and Jean Morgensen. He entered North Carolina State University as an undergraduate in the fall of 2001 and graduated with a B.S. in Electrical Engineering. He continued his Electrical Engineering studies by enrolling in the graduate program at NCSU beginning fall of 2005 and obtained the M.S. degree in 2008. He has greatly benefitted from internships with two different companies: Marvell Semiconductor in 2004, 2006 and RF Micro Devices in 2007, 2009.
ACKNOWLEDGEMENTS

I would like to thank Dr. Lunardi for agreeing to act as Committee Chair throughout my Ph.D. research. I greatly benefited from your guidance in technical writing and willingness to offer suggestions for challenging problems. I am also thankful for the encouragement and advising given by Dr. Barlage that began when I was an undergraduate and continued through completion of my Ph.D. Thank you for giving me the opportunity to attend graduate school, something I never had any intention of doing full-time. Thank you to Dr. Johnson and to Dr. Bilbro for serving on my advisory committee and for the useful discussions your questions generated.

I would also like to acknowledge the contributions from my past and present colleagues. Dr. Jaehoon Park, you were an excellent addition to our group and thank you for the many conversations regarding semiconductor physics. Merve Ozbek, thank you for your willingness to assist with simulation questions and also for your support on the day of my defense. Dr. Matt Veety, thank you for all the hard work devoted towards fabricating devices. It’s been quite an experience since our undergraduate days, and I enjoyed the many technical discussions over the years.

I am indebted to Dr. Bongmook Lee who graciously agreed to collaborate and share data from his dissertation devices. Without your data I would never have noticed some crucial model errors nor improved the procedures for parameter extraction. Thank you to Dr. Misra for allowing me access to the measurement lab over the years, and thank you to Dr. Steven Novak for patiently answering measurement questions.
# TABLE OF CONTENTS

List of Tables ................................. vi

List of Figures ................................. vii

Chapter 1  Introduction ......................... 1
  1.1 Overview of contents ......................... 4
     1.1.1 Optimization details ...................... 6
     1.1.2 Comments on simulated curves .............. 6

Chapter 2  Charges in the metal-oxide-semiconductor system ............... 8
  2.1 Semiconductor charge ....................... 10
  2.2 The surface potential equation ............... 15
  2.3 Quantum mechanical effects ................... 18
     2.3.1 Modifications for inversion in p-channel MOS devices .... 23
     2.3.2 Summary of useful charge density expressions ............ 24
  2.4 Interface trapped charge ..................... 25
     2.4.1 Trap occupation ............................ 26
     2.4.2 Interface trapped charge density ............... 28
     2.4.3 Approximating trapped charge density ............... 33
     2.4.4 Expanding $Q_{it}'$ ......................... 38

Chapter 3  Drain current ......................... 42
  3.1 Charge Sheet Approximation details ............ 44
  3.2 Behavior of trap-affected $Q_i'$ ............... 49
  3.3 Effective Mobility ............................. 52
  3.4 Calculating drain current ...................... 54
     3.4.1 Effects of $D_{it}$ on DC drain current ............ 61
     3.4.2 Velocity saturation ........................... 66
     3.4.3 Series resistances ............................ 67
     3.4.4 Current saturation ............................ 71
  3.5 Model Summary ................................. 74

Chapter 4  Mobility analysis ...................... 76
  4.1 Discussion of the mobility measurement from C-V ............ 77
     4.1.1 Analytical correction for the trap-free case ............ 82
     4.1.2 Accounting for interface states ................... 94
  4.2 Mobility from drain current .................... 96
     4.2.1 Application to simulated and real trap-free data .......... 101
Chapter 5  Analysis of the insulator-semiconductor interface through drain current 110
5.1 Doping investigation from capacitance-voltage ................................. 113
5.1.1 Extracting effective doping from low frequency C-V measurements .. 117
5.2 Drain current based $D_{it}$ analysis .................................................... 119
5.2.1 Method 1 ..................................................................................... 120
5.2.2 Method 2: arbitrary $t_{ox}$ ............................................................... 126
5.3 Application of methods to silicon MOSFETs ................................. 132

Chapter 6  Characterization of gallium nitride MISFETs ..................... 139
6.1 GaN-specific modeling concerns ....................................................... 140
6.2 Analysis of a three-terminal GaN MISFET ................................. 144
6.2.1 Supporting theory ........................................................................ 144
6.2.2 Measurement results ................................................................. 150
6.3 Measurement of a four-terminal GaN MISFET .......................... 157

Chapter 7  Conclusions and suggested work ..................................... 163
7.1 Suggested work ................................................................. 165

References ................................................................. 167
LIST OF TABLES

Table 5.1  nMOS parameter summary. .................................................. 135
Table 5.2  pMOS parameter summary. .................................................. 135
Table 6.1  Summary of results obtained from Method 2. ....................... 153
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Simple MOSFET cross-section (without gate overlap) showing chosen positioning of axes.</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Total semiconductor charge density with different values for $V$, defined in the paragraph above equation (2.10).</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>Plots of surface potential versus gate voltage with varying (a) oxide thickness, (b) doping, (c) imref splitting. $V_{FB} = 0V$ for all curves and $V = 0$ (unless otherwise indicated).</td>
<td>17</td>
</tr>
<tr>
<td>2.4</td>
<td>C-V plots for different oxide thickness values (a) 2.5nm, (b) 25nm, (c) 50nm, showing the discrepancy between the classical and QM approaches.</td>
<td>22</td>
</tr>
<tr>
<td>2.5</td>
<td>Filled/empty states and band bending.</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>Difference in occupation functions where $F$ is the exact steady-state version and $F_{FD}$ is the simplified quasi Fermi-Dirac version. The exact version indicates that $V$ can only push the falling edge leftward until around midgap, at which point the simplified version becomes incorrect.</td>
<td>27</td>
</tr>
<tr>
<td>2.7</td>
<td>Effect of bandtail parameter $b$ on interface state charge density. Without the $D_c$ term $Q_{it}$ would increase linearly with respect to $\psi$, and then saturate as $E_c$ is bent down towards $E_F$. $V = 0$.</td>
<td>31</td>
</tr>
<tr>
<td>2.8</td>
<td>Effect of different densities on interface state charge density for $V = 0$.</td>
<td>31</td>
</tr>
<tr>
<td>2.9</td>
<td>Effect of different densities on the surface potential equation for $t_{ox} = 30$nm. In (b), solid line: $D_i = 2 \times 10^{12}$, $D_c = 5 \times 10^{13}$; dashed line: $D_i = 10^{12}$, $D_c = 10^{13}$; dot-dashed line: $D_i = 5 \times 10^{11}$, $D_c = 10^{12}$.</td>
<td>32</td>
</tr>
<tr>
<td>2.10</td>
<td>Quality of MID. Solid line: developed approximation; closed circles: exact result. Inset shows percent error for $V = 0.5$ since for $V = 0$ the result is exact. The percent error is the same for other values of $V$.</td>
<td>34</td>
</tr>
<tr>
<td>2.11</td>
<td>Magnitude of error associated with the summation and $\sigma$ integration approach for $V = 0$. As the ratio $D_i/D_c$ is increased the error will decrease. $V = 0$ and $T = 25C$. All of the error is from CON because MID is exact at $V = 0$.</td>
<td>36</td>
</tr>
<tr>
<td>2.12</td>
<td>Magnitude of error associated with the summation and $\sigma$ integration approach for $V = 0.5$. As the ratio $D_i/D_c$ is increased the error will decrease.</td>
<td>37</td>
</tr>
<tr>
<td>2.13</td>
<td>Delta in predicted gate-bulk voltage by evaluating SPE at $t_{ox} = 30$nm. Increasing $V$ shifts the curves toward the right and highlights error caused by MID.</td>
<td>37</td>
</tr>
</tbody>
</table>
Figure 2.14 First derivative of $Q'_i$ for $V = 0$ (MOS capacitor case). In (b) each solid exponential line is a plot of $D_i + D_c e^{(\psi_F + E_F - E_c)/b}$ to investigate the common approximation $C_i / q \approx D_i$. The x-axis can be thought of as $E - E_v$ similar to Figure 2.7.

Figure 3.1 Surface potential swept from $\phi_F$ to $3\phi_F$ (an extreme value but chosen to show the decrease in error). In (a), the largest curve corresponds to the greatest doping and vice-versa.

Figure 3.2 Percent error of the drift-diffusion charge sheet model for $I_{ds}$ relative to the numerically evaluated Pao-Sah expression. The x-axis is different in (b) because of voltage stretchout by $D_i$.

Figure 3.3 Legend indicates value of $\psi_{s0}$ (source-side surface potential). (a) demonstration of linearity when traps are absent. (b)-(d) traps cause non-linearity until a large enough gate bias is applied (strong inversion triode region). $V_{DS} = 0.5V$, $t_{ox} = 30nm$, $N_A = 10^{17}cm^{-3}$. $Q_{i0,L} = Q_i(\psi_{s0,L}, V_{SB, DB})$.

Figure 3.4 Effect of field-dependent mobility on channel inversion charge. $r$ equals $\mu_e$ evaluated at the channel midpoint, which clearly produces a convenient alternative to integrating $\mu Q'_i$. The mobility-affected quantities were normalized by $\mu_0 = 350cm^2/Vs$; also, $K = 6(10^{14})V/s$. Gate bias chosen to operate simulated device in the triode region.

Figure 3.5 Percent error of expression (3.19) relative to discretely integrated (3.11). Lines without markers are for $V_{DS} = 100mV$ and lines with markers are for $V_{DS} = 1V$. Higher $V_{DS}$ increases error slightly at larger band-bending. Calculated for special case of unity $\mu_e W L$ to demonstrate quality of approximation to $\int Q'_i d\psi_s$.

Figure 3.6 Percent error of expression (3.20) relative to discretely integrated (3.11) for $V_{DS} = 100mV$.

Figure 3.7 Effect of increasing $D_c$ with $\mu_e W L = 1 cm^2/V \cdot s$, $N_A = 10^{17}$, $b = 50 meV$, $D_i = 5(10^{11})$. The x-axis is shifted for each curve by its initial value. $V_{GS}$ for each curve computed by starting and ending at source-side surface potentials $\phi_F$ and 1.0V.

Figure 3.8 Effect of increasing $b$ with $\mu_e W L = 1 cm^2/V \cdot s$, $N_A = 10^{17}$, $D_c = 3(10^{13})$, $D_i = 5(10^{11})$. The x-axis is shifted for each curve by its initial value. $V_{GS}$ for each curve computed by starting and ending at source-side surface potentials $\phi_F$ and 1.0V.

Figure 3.9 Increasing values of $t_{ox}$ reduce oxide capacitance density and strengthen the impact of $D_i$.

Figure 3.10 Velocity-field curves for three materials (references in text).
Figure 3.11 Characteristics of a simulated device with $t_{ox} = 10\text{nm}$, $N_A = 10^{17}$, $D_i = 0$, $D_c = 10^{13}$, $b = 0.05$, $\mu_e = \mu_0 = 150$. $V_{GS}$ from top to bottom: [3.6, 4, 4.6] V. .................................................... 70

Figure 3.12 Characteristics of a simulated device with $t_{ox} = 10\text{nm}$, $N_A = 10^{17}$, $D_i = 5(10^{11})$, $D_c = 3(10^{13})$, $b = 0.05$, $\mu_e = \mu_0 = 150$, $V_{DS} = 0.1\text{V}$. ................................ 70

Figure 3.13 Top curve to bottom: $L = 100$, 10, 1 um. Markers: numerical search; line: analytical expression. $N_A = 10^{17}\text{cm}^{-3}$, $t_{ox} = 30\text{ nm}$, $\mu_e = \mu_0 = 100$. Threshold is near 2.7V. ............................... 73

Figure 4.1 Actual mobility and equation (4.5) applied at three different drain biases. All curves tend towards the same limiting value at high gate bias but include some artificial roll-off near threshold (around 0.6V). $W = L$, $N_A = 10^{16}\text{cm}^{-3}$, $t_{ox} = 200\text{Å}$, $T = 25\text{C}$, trap-free (same parameters used for any other fictional device results to follow). ......................... 81

Figure 4.2 Quality of $\Delta$ for two small $V_{DS}$ values as function of $V_{GS}$. (c) indicates threshold. ............................................................... 85

Figure 4.3 Effective mobility at $V_{DS} = 100\text{ mV}$. .................................................. 89

Figure 4.4 Effective mobility at $V_{DS} = 50\text{ mV}$. .................................................. 90

Figure 4.5 Effective mobility at $V_{DS} = 10\text{ mV}$. .................................................. 91

Figure 4.6 All data obtained from C.L. Huang (reference in text). The analytical correction works well since doping and oxide thickness were well-known and $D_{it}$ was very low. ........................................ 93

Figure 4.7 Drain current reduced by series resistance and field-dependent mobility. $R_{sd} = 5\text{kΩ}$. ................................................................. 98

Figure 4.8 Dotted line indicates the linear fit. (a) estimates $\mu_0$ as 102, compared to 100 actual. (b) estimates $\theta_2$ as $6.99\cdot(10^{-3})\text{V}^{-2}$, compared to $6.94\cdot(10^{-3})\text{V}^{-2}$ actual. (c) estimates $\theta_1^e$ as $9.55\cdot(10^{-2})\text{V}^{-1}$, compared to $9.67\cdot(10^{-2})\text{V}^{-1}$ actual. (d) is the fit after direct extraction with equation (4.31). .................................................. 102

Figure 4.9 The digitized data from C.L. Huang is indicated with markers (reference given in text). The slope of (a) is related to $\mu_0$ and the slope of (b) is $\theta_1$. In (c) the fit using equation (4.21) is compared against the measured data. Because of the noise in (a) at $V_{DS} = 20\text{mV}$ the low field mobility is overestimated, worsening the result in (c). ................. 104

Figure 4.10 Behavior of transconductance and subparts of $Q_i'$ for a trap-free device. Parameters of the simulated device: $t_{ox} = 200\text{Å}$, $N_A = 10^{16}\text{cm}^{-3}$, $\mu = 100/\left(1 + E_c^0/E_0^2\right)$, $E_0 = 10^6\text{MV/cm}$, $V_{FB} = -0.5\text{V}$. ......................................... 107

Figure 4.11 Behavior of transconductance and subparts of $Q_i'$ for a constant $D_{it}$: .......................... 107

Figure 4.12 Behavior of transconductance and subparts of $Q_i'$ for a constant $D_{it}$ greater than $C_{ox}'/q$. ........................................ 108
Figure 6.3 Voltage stretchout and frequency dependence related to interface states. Top dotted line is related to $C'_g(f \rightarrow 0)$; bottom dotted line is $C'_g(f \rightarrow \infty)$. Solid lines from left to right: 1kHz, 31.6kHz, 1MHz. The title in each plot gives the interface state density relative to the oxide capacitance density. (b) is an extreme case where greater than 15V of stretchout occurs.

Figure 6.4 Capacitance and conductance results over a range of frequencies. (c) is the corrected curve after deembedding oxide capacitance density, and (d) is the rescaled version showing the peak moving downward in frequency. The dotted line in (a) indicates the estimated oxide capacitance.

Figure 6.5 Representative GaN MISFET that could survive out to very high gate voltages. From C-V, SiO$_2$ thickness is around 70 nm. The fluctuation in (c) is the result of discrete differentiation, which has “amplified” some minor deviation in (a). The minor deviation in (a) could be attributed to the measurement equipment, poor probe contact to pads, or some small change in the interface properties due to biasing.

Figure 6.6 Outcome of Method 2 applied to the data in Figure 6.5. In (a) the markers represent measured data, solid lines are the fit. In (b) both the lines and markers represent the resulting mobility. $W = L = 64$um and $t_{ox} \approx 70$nm.

Figure 6.7 Reconstructed ideal case for device of Figure 6.5 where $D_{it} = 0$.

Figure 6.8 Real and imaginary components of the gate small-signal response. 20 frequencies evenly spaced logarithmically from 1kHz (left-most curve) to 1MHz (right-most curve). Series resistances reduce the peak capacitance at high frequencies. Representative plot of typical device.

Figure 6.9 The family of curves are tilted upward by the presence of junction leakage. Representative plot of typical device.

Figure 6.10 Log and linear scale plots of drain current showing how $I_{on}$ to $I_{off}$ is not much greater than a decade. Representative plot of typical device.

Figure 6.11 One quadrant (repeated in three other quadrants) in a 1 cm x 1 cm square. The blue layer is final metal layer used for probing and understandably dominates the layout. Red indicates ion-implant while grey corresponds to mesa isolation. Not visible due to final metal layer are the n-ohmic window, p-ohmic window, and via.
Chapter 1

Introduction

The introduction of alternative insulator and substrate materials for metal insulator semiconductor field-effect transistor (MISFET) fabrication has the potential to create non-idealities in the insulator-semiconductor interface properties, an area greatly improved for the silicon metal oxide semiconductor field-effect transistor (MOSFET) over the past five decades [1]. These non-idealities collectively violate the basic assumption on which most electrical MOS-FET models are built, namely that the only source of bias-dependent charge is from the semiconductor itself. A model that incorporates interface non-idealities could therefore be useful and flexible in gauging improvement in fabricated devices.

The insulator-semiconductor interface is complex because of the formation of interface states, often referred to as surface states or interface traps. Lattice periodicity is interrupted in the transition from the semiconductor to the insulator leaving some bonds unsatisfied or “dangling.” The Disorder Induced Gap States model considers the existence of a disordered layer between the insulator and semiconductor as the source of interface states [2]. Under the DIGS theory this layer produces a continuous distribution of occupiable electron states throughout the bandgap of the semiconductor. The density of these interface states can vary
substantially depending on the quality of the surface passivation and how effective it is in limiting the trapping process [3]. The interface state density (defined as $D_{it}$) as a function of energy can be described empirically with a pair of exponentials and is often described as U-like [4]. The $D_{it}$ distribution has alternatively been modeled using sums of Gaussians [5] that could resemble a U-like distribution depending on the Gaussian parameters.

In a MOSFET the applied gate bias directly controls the charges present underneath the gate electrode. By charge conservation, the sum of all charges below the gate equals the total gate charge. Interface states if present will contribute a charge density that varies with band-bending along with the semiconductor charge density. As gate bias changes, both interface state charge density and semiconductor charge density will respond in balancing the gate charge. Therefore, at a given gate bias the semiconductor charge density varies from the value without interface states. From an electrical measurement point of view this behavior is termed gate voltage stretch-out.

Voltage stretch-out absent in optimized devices is not usually included in models for integrated circuit simulation. However, its inclusion would allow for better understanding between measured and simulated electrical data of non-optimized devices and device parameter extraction. In this work the electrical data of interest is MOSFET drain current. Including stretch-out in drain current has been attempted over localized gate bias regions by considering simplifications of the expression for $D_{it}$. The complexity of exact inclusion lies in the definition of the interface trapped charge expression, written as an integral over the bandgap of density of states with an occupation probability [6]. While the inclusion of interface states is important for correctly describing drain current from weak through strong inversion, the necessary integral for a U-like description is not easily solved or approximated.

The most common approximation for interface state density is through a constant value over the bandgap. In this case the charge density integral can be solved and rapidly calculated
over all gate biases. P. Muls used this simplification to develop closed-form approximations for drain current in weak inversion [7]. Extending the result outside weak inversion is feasible but would involve some degree of approximation for the integration of inversion charge density to determine drain current [8, 9]. Unless a U-like distribution is considered, a constant interface state density cannot explain the slow transition between linear region drain current and that in subthreshold. It would be more accurate to consider an energy-dependent $D_{it}$ function for real devices when modeling drain current from subthreshold through strong inversion.

Common approaches that include the distribution over energy for calculating drain current are either simplified or completely numerical. These approaches can be found in the domain of silicon carbide MOSFETs and were deemed necessary in order to gain theoretical understanding of measured device behavior since SiC MOSFETs have been found to suffer from very high interface state density [10]. The most common simplification for calculating charge density with non-constant $D_{it}$ is to assume the probability occupation function with a sharp transition thus changing the limits of integration [11, 12, 13]. This approach works to a certain degree but fails in strong inversion due to the lack of the natural saturation of charge density as the minority carrier band edge bends toward the Fermi level. The simplification is acceptable only if the interface state charge density is negligible relative to majority or minority carrier charge density when the simplification fails.

Exact inclusion of U-like $D_{it}$ has been performed by S. Potbhare in a numerical device simulator [14]. The exponentially-varying behavior of $D_{it}$ near the band edges was attributed to “bandtail” states, with electrical characteristics obtained by discretization of the two-dimensional Poisson’s equation. The U-like distribution was described empirically as two exponentials summed with a constant, giving five coefficients. Of these five, it is often the case that only two are needed to model MOSFET drain current for the case of large $D_{it}$ near the minority carrier band edge.
1.1 Overview of contents

This present work was first directed towards developing a compact model for drain current that completely accounted for interface states. The U-like distribution in [14] was chosen to describe interface states. A compact drain current result permits the rapid evaluation of drain current without the long simulation times inherent to two-dimensional numerical analyses. The compact modeling goal is to develop an expression for drain current not rooted in regional approximations of interface state charge density, a topic not usually covered in the literature. This subject is partially covered in Chapter 2 that begins with a discussion of semiconductor and interface state charge densities following the classical one-dimensional solution of Poisson’s equation for the long-channel MOSFET.

The approximations to the interface state charge density integral would only be of sufficient precision for known U-like trap coefficients. For the case of arbitrary trap coefficients it is found that numerical integration provides a better choice, with an effort to transform the integral and to reduce the number of points used in integration. In addition the common triangular well approximation to the quantum-mechanical treatment of minority carriers is discussed in anticipation of further device modeling of thin oxide silicon MOSFETs. Here, the quantum mechanical treatment has been slightly modified for the presence of interface states by inclusion of their effects on the surface field.

Chapter 3 continues towards the compact drain current expression and is based upon the charge density results obtained in the previous chapter. The resulting expression’s accuracy indicates some amount of drain bias dependence and consumes at least twice as much computation time as a trap-free model. A simplified result has been written for low drain-source bias, outcome of applying Simpson’s rule for the case of two subintervals. Second order effects such as velocity saturation and series resistances have been included. A pseudo-analytical
expression for the drain-source saturation voltage has also been developed to support channel length modulation. These second order effects were implemented following the approach of the surface potential based models PSP [15] and MM11 [16].

The channel mobility has been left in a general form during development of the compact drain current expression. However, mobility measurements could be used as a means of empirically modeling its dependence on field. In Chapter 4 the trap-free case is considered as a means of rigorously investigating the details of effective mobility measurement. The origin of the “drain bias discrepancy” is discussed with a newly proposed analytical correction, an important tool needed for comparison of device results and so that the effects of Coulomb scattering are not overestimated. Limitations imposed by interface states are also investigated as preparation for analysis of real devices.

Chapter 5 considers alternatives to $D_{it}$ estimation solely using drain current. This chapter was shaped by earlier work with fabricated GaN MISFETs lacking a bulk contact, which limited the experimental measurements. Two methods are presented with the first one capable of obtaining estimates later used for optimization. The second is capable of estimating trap parameters, surface doping, flatband voltage, and mobility through intelligent manipulation of the measured drain current characteristic. Both methods are then applied to experimental data of silicon-based MOSFETs fabricated with thin gate insulators and alternative dielectrics. The role of non-uniform doping is also investigated by the one-dimensional finite difference method to understand anomalous depletion region capacitance-voltage behavior.

Chapter 6 uses the results of the previous chapters to investigate GaN-based MISFETs fabricated at NC State University using the methods presented in the earlier chapters. These devices suffer from very high interface state density and are representative of a typical non-optimized device. One of the methods from Chapter 5 is applied to extract the trap parameters, of which only those corresponding to the minority carrier band edge are needed due to junction
leakage. Finally, Chapter 6 concludes with measured results from a second generation of GaN-based MISFETs in a new design.

1.1.1 Optimization details

Optimization is used intensely in Method 2 presented in Chapter 5 because of the absence of analytical expressions describing the drain current from subthreshold through strong inversion. For this work, optimization with non-linear least squares in MATLAB was employed to fit measured quantities [17]. Given a set of coefficients, the non-linear least squares routine minimizes the sum of squares composed of the difference between the measured and simulated quantity, typically involving some manipulation of drain current. Upper and lower bounds were given for each parameter with appropriate starting values often estimated from another technique. Optimization was generally allowed to continue until convergence on either the parameter values or sum of squares was met within a maximum number of iterations.

1.1.2 Comments on simulated curves

Using the derived equations contained entirely within the chapters of this thesis, several simulated plots have been generated for demonstration. No comparison to any transistor computer aided design (TCAD) or any simulation software has been included. The surface potential equation (SPE) in equation (2.14) is the essential foundation on which all other equations are evaluated, since it relates applied gate voltage and semiconductor surface potential. The SPE is solved numerically, often by the simple bisection method for convenience without any concern on the computational speed.

All plots are generated at room temperature using known constants such as permittivity, Boltzmann’s constant, Planck’s constant, free electron mass, and elementary charge. Variables
representing charge densities are marked with a “prime” superscript and are the same as charge per unit area. Material constants used during simulation were theoretical values with [18]. For Chapter 5 the silicon value of intrinsic carrier concentration that agrees with experiment was chosen, $n_i \approx 10^{10} \text{cm}^{-3}$ [19]. Further, the intrinsic energy level was taken to be at midgap for Chapter 5.
Chapter 2

Charges in the metal-oxide-semiconductor system

Analysis of a field effect transistor is inherently a three-dimensional problem described by the coupled Schrodinger and Poisson equations [20]. For thick oxides quantization effects are ignored and only Poisson’s equation is considered, which is the “classical” approach. This is a highly desirable approach since self-consistent solution of Schrodinger’s and Poisson’s equations is a slow, numerical analysis. In contrast, the classical approach is capable of producing compact expressions for drain current that can be evaluated rapidly.

A simple MOSFET cross-section without gate overlap is shown in Figure 2.1 with z-axis pointing into the device, perpendicular to the x- and y-axes. Ignoring edge effects, charge distribution can be assumed identical along the z-axis and analysis of Poisson’s equation reduces to a two-dimensional problem. Even with this reduction only a numerical solution is possible without additional simplification because of the dependence $\rho$ has on potential, $\psi$. An accurate, analytical, closed-form solution of Poisson’s equation is instead obtained from a specialized one-dimensional case. This specialized case is based on the Gradual Channel
Figure 2.1: Simple MOSFET cross-section (without gate overlap) showing chosen positioning of axes.

Approximation where it is assumed that the horizontal field is negligible relative to the vertical field [21]. The one-dimensional case is technically a MOS capacitor, where the potential extending vertically from the gate to bulk is assumed uniform at any other \((x, z)\) coordinate underneath the gate.

\[
\frac{d^2 \psi(y)}{dy^2} = -\frac{\rho(\psi)}{\varepsilon_s}
\]

(2.1)

Treatment in two [22] or even three [23, 24] dimensions is possible for highly simplified subthreshold cases, a topic popular when transistor scaling caused threshold-referenced models to become less and less accurate. Such methods rely on setting appropriate boundary conditions and using separation of variables in order to find a solution. The results are generally used to provide a \(\Delta V_{th}\) and are unfortunately not accurate for a model covering all possible gate biases.

This chapter begins by briefly discussing some of the details in the classical one-dimensional solution of Poisson’s equation culminating in a description of total semiconductor charge den-
sity (charge per unit area), \( Q_s' \). With \( Q_s' \) known the classical treatment is followed to demonstrate the relationship between gate voltage and surface potential. Interface states are then discussed in regards to the relationship between state density and state charge density. The density to charge density relationship requires a non-trivial solution if interface state density is not constant with respect to energy and necessary approximations are developed. Finally, a compact trap-affected drain current model is presented that is non-regional and properly models the subthreshold to strong inversion transition.

## 2.1 Semiconductor charge

The charge density (charge per unit volume), \( \rho \), is composed of free electrons, free holes, ionized donors, and ionized acceptors. When a gate voltage is applied to a MOSFET a potential is induced within the semiconductor substrate. This potential, labeled as \( \psi \), is a consequence of charge balance between the applied gate charge and induced substrate charge. The induced substrate charge’s magnitude is some combination of the four charge types described below.

\[
\rho = q(p - n + N_D^+ - N_A^-)
\]  

(2.2)

The free carrier concentrations are described by Fermi-Dirac statistics of order one-half. \( N_{c,v} \) are the effective electron and hole density of states, \( \psi \) is semiconductor band-bending, \( E_F \) is the bulk Fermi level, and \( E_c \) is the conduction band energy level.
\[ p = N_v \frac{F_{1/2}(\eta_v)}{\Gamma_{3/2}}, \text{ with } \eta_v = \frac{-q\psi + E_v - E_F}{kT} \]

\[ n = N_c \frac{F_{1/2}(\eta_c)}{\Gamma_{3/2}}, \text{ with } \eta_c = \frac{q\psi - (E_c - E_F)}{kT} \]  

Equations (2.3) effectively reduce to Maxwell-Boltzmann statistics for values of \( q\psi + E_F \) several \( kT \) away from the conduction band edge [25]. These reduced equations are usually rearranged in terms of the intrinsic carrier level by replacing \( n \) or \( p \) with \( n_i \) and \( E_F \) with \( E_i \).

\[ n \approx N_c e^{(q\psi - (E_c - E_F))/kT} \]

\[ p \approx N_v e^{(-q\psi + E_v - E_F)/kT} \]  

Carrier degeneracy is important in strong inversion when the minority carrier band edge has crossed the Fermi level. At lower temperatures the effect is more pronounced [26]. Every substrate material is slightly different in regards to degeneracy effects because of differences in effective conduction/valence band density of states. While easily accomplished, developing a compact model with Fermi-Dirac statistics is usually not necessary unless low temperature capability is desired. Degeneracy can be ignored and the mobility-field relationship modified to correctly describe drain current.

Electrons and holes in excess of the intrinsic density are added to a semiconductor through the addition of impurities. Donor-like atoms donate an electron to the conduction band while acceptor atoms accept an electron from the valence band. The donation or acceptance of an electron ionizes the immobile impurity atom to give it charge. While complete ionization at room temperature is often assumed, this is only true for dopants with a small ionization energy in relation to the conduction or valence band. Band-bending will eventually ionize
even dopants with very large ionization energies, which effectively increases threshold slightly. Using [27] to include degeneracy factors $g_D, g_A$, the ionized levels can be written as a function of band bending.

\[
N_D^+ = N_D \left[ 1 + g_D e^{(q\psi + E_F - E_D)/kT} \right]^{-1}
\]

\[
N_A^- = N_A \left[ 1 + g_A e^{(-q\psi + E_A - E_F)/kT} \right]^{-1}
\]

The donor and acceptor energies, $E_A, E_D$, depend on the substrate material and dopant considered. Assuming complete ionization at room temperature is usually acceptable.

The Fermi level, $E_F$, that has appeared in each equation resides anywhere within the bandgap and determines the equilibrium carrier concentration. It acts as the reference level for all cases of band-bending when bias is applied. In general, one finds it by setting $\rho = 0$ and numerically solving for $E_F$ (relative to another energy level) with $\psi = 0$. Alternatively, a simple result is found from (2.4) for a single, completely ionized dopant with concentration much greater than the intrinsic concentration. $\phi_t$ is the thermal voltage given by $kT/q$.

\[
\phi_F = \frac{E_i - E_F}{q}
\]

\[
= \phi_t \ln \left( \frac{N_A}{n_i} \right) \text{p-type}
\]

\[
= \phi_t \ln \left( \frac{n_i}{N_D} \right) \text{n-type}
\]

The Gradual Channel Approximation modifies the minority carrier concentration expression to account for carrier injection by the source. The application of a drain-source bias causes
the minority carrier quasi-Fermi level to split away from the bulk Fermi level. The difference between the bulk Fermi level and the quasi-Fermi level is written as \( qV \), which after normalizing by \( q \) can be named the channel-bulk potential [28] or more formally as imref splitting [15]. It can be treated as a parameter that takes on values \( V_{SB} \) at the source and \( V_{DB} \) at the drain. For an n-channel MOSFET equations (2.4) are re-written in terms of \( N_A \) and the electron density is modified by \( V \).

\[
\begin{align*}
    n &= N_A e^{\frac{(\psi - 2\phi_F - V)}{\phi_t}} \\
    p &= N_A e^{-\frac{\psi}{\phi_t}}
\end{align*}
\]  

(2.10)

The one-dimensional analysis of Poisson’s equation can be generally described as follows. One begins by multiplying both sides of (2.1) by \( 2 \frac{d\psi}{dy} \) and recognizing that the LHS equals \( dE^2/dy \). The resulting equation can be integrated from deep in the bulk (\( E(\infty) = 0, \psi(\infty) = 0 \)) to an arbitrary point \( y \) after multiplying both sides by \( dy \), thus giving the square of the vertical electric field. Afterwards, the total semiconductor charge density (charge per unit area), \( Q'_s \), can be calculated using the integral form of Gauss’s Law in one dimension evaluated from the surface (\( y = 0 \)) to deep in the bulk where the electric field is zero. The value of \( \psi \) at \( y = 0 \) is termed the surface potential and labeled \( \psi_s \). The relationship between semiconductor charge density and surface field is,

\[
E_s = -\frac{Q'_s}{\varepsilon_s}.
\]  

(2.11)

The total semiconductor charge has units \( C/cm^2 \) and can be thought of as normalized to \( WL \) (at least for a Metal-Oxide-Semiconductor capacitor). The value of total charge is negative.
Figure 2.2: Total semiconductor charge density with different values for $V$, defined in the paragraph above equation (2.10).

for $\psi_s > 0$ and positive otherwise. The form of $Q'_s$ written below is taken from [16]. Slightly different forms can be obtained that produce imaginary results around flatband [29, 30].

$$Q'_s = \mp \sqrt{2qN_A\varepsilon_s} \left[ \psi_s + \phi_t \left( e^{-\psi_s/\phi_t} - 1 \right) + \phi_t e^{-\left( V + 2\phi_F \right)/\phi_t} \left( e^{\psi_s/\phi_t} - 1 \right) \right]^{1/2}$$

The absolute value of equation (2.12) is plotted in Figure 2.2. For $\psi_s < 0$ the surface is accumulated and charge density increases exponentially. Raising $\psi_s$ above zero introduces depletion until sufficient band-bending inverts the surface, increasing $Q'_s$ exponentially once again.
2.2 The surface potential equation

While an explicit expression for the total semiconductor charge has been found, it is in terms of surface potential, not the applied gate bias. To relate band-bending with gate bias one writes an equation balancing all charges, an equation balancing all potentials, and applies Gauss’s Law to an ideal charge-free oxide [28]. Treating the oxide as charge-free is the same as assuming all oxide charge appears at the semiconductor interface.

\[ Q_G' + Q_s' + Q_{ox}' = 0 \]

\[ V_{GB} = \phi_{ms} + \psi_s + \psi_{ox} \]  \hspace{1cm} (2.13)

\[ Q_G' = C_{ox}' \psi_{ox} \]

The charges to be balanced include the gate charge, semiconductor charge, and oxide charge. The oxide charge contains contributions from interface states and any fixed/trapped charge, labeled as \( Q_{it}' \) and \( Q_o' \). Interface states are bias-dependent and have a value at flatband denoted \( Q_{it00} \), which is the interface state charge density evaluated at zero band-bending and zero imref splitting. Combination of all equations results in a transcendental equation.

\[ V_{GB} = V_{FB} + \psi_s - \frac{Q_{it}'(\psi_s) - Q_{it00}}{C_{ox}'} - \frac{Q_s'}{C_{ox}'} \]  \hspace{1cm} (2.14)

Equation (2.14) is the “surface potential equation” (SPE) following the terminology of [15]. \( V_{GB} \) can be easily calculated for a given surface potential but the reverse is generally desired. To find \( \psi_s \) for a given \( V_{GB} \), one can choose to solve numerically or resort to an analytical function [31].
The flatband voltage describes the voltage at which $\psi_s = 0$ and there is no net semiconductor charge. Composed of oxide charge and the metal-semiconductor work function, it is the condition for flat energy bands and equilibrium-like conditions in the substrate. The metal-semiconductor work function can be calculated assuming a metal gate with known work function $\phi_m$. The semiconductor work function is the energy required to move an electron from the Fermi level to vacuum. Easily visualized by drawing a band diagram, it is expressed in terms of the substrate’s electron affinity, band gap, and Fermi level.

$$V_{FB} = \phi_{ms} - \frac{Q_o'}{C_{ox}} - \frac{Q_{it00}}{C_{ox}}$$  \hspace{1cm} (2.15)

The SPE maps gate voltage to band-bending, and a unique mapping exists for every value of oxide thickness and doping. Interface states also distort this mapping and delay the onset of inversion by balancing gate charge along with semiconductor charge. Ignoring contributions from $Q_{it}'$ some $\psi_s$ versus $V_{GB}$ curves can be generated. $\psi_s(V_{GB})$ is plotted for different oxide thicknesses and doping in Figures 2.3a and 2.3b. Surface potential was calculated from $-0.1V$ to $2\phi_F + 5\phi_i$ for each curve. The surface potential increases almost linearly upward from flatband until inversion is reached, at which point $\psi_s$ begins to saturate. This occurs because the inversion charge density increases rapidly and dominates the SPE. For the same bulk doping concentration different oxide thicknesses delay threshold as shown in 2.3(a). In 2.3(b), larger p-type doping moves the bulk Fermi level towards the valence band, requiring greater band bending to reach inversion.

The effects of different values of $V$ are shown in Figure 2.3c. In the special case that parameter $V$ equals $V_{DB}$, then Figure 2.3c shows that larger drain-bulk biases decrease inversion at the drain end of the channel.
$N_A = 10^{17} \text{ cm}^{-3}$

$t_{ox} = 30 \text{ nm}$

$N_A = 10^{17} \text{ cm}^{-3}, t_{ox} = 30 \text{ nm}$

Figure 2.3: Plots of surface potential versus gate voltage with varying (a) oxide thickness, (b) doping, (c) imref splitting. $V_{FB} = 0\text{V}$ for all curves and $V = 0$ (unless otherwise indicated).
2.3 Quantum mechanical effects

Band-bending at the insulator-semiconductor surface caused by an applied gate bias confines free carriers in a potential well at discrete energy levels above the conduction band. The density and spatial distribution of free carriers are determined by numerical analysis of the coupled Schrodinger and Poisson equations, a computationally lengthy process unsuitable for a compact model. Quantum confinement requires more band-bending to achieve the same inversion charge density, a consequence of the lowest allowed energy level being some finite energy above $E_c$. Also, the distance from the interface to the centroid of the minority carrier distribution is slightly different from the classical result, giving an effective increase in the oxide capacitance density. The effects of quantum confinement are readily observed for devices with thin oxides and/or high doping: on a plot of gate capacitance versus voltage the peak capacitance fails to reach $C'_{ox}$ and the threshold’s magnitude increases [32].

Approximations for the lowest populated subband’s energy level can be found by either treating the potential well as triangular or assuming only the lowest subband is filled [20]. These approximations can be used to create an energy offset relative to the conduction band. The van Dort model considers quantum confinement as an effective bias-dependent bandgap widening term [33]. The intrinsic carrier concentration is replaced with a quantum mechanical version with energy gap offset, $\Delta E$, proportional to the surface field raised to the two-thirds power. However, the effective mass used in the computation of $\Delta E$ loses some physical meaning since more than one subband is involved [34]. A more physical approach would include more than one subband.

This work considers only the quantum confinement in the inversion region since predicting drain current is of foremost importance. In accumulation it has been shown that quantum mechanical treatment of majority carriers is necessary to calculate gate capacitance. The tri-
angular well approximation has been found to be useful as a means of developing a bandgap widening term [35]. However, implementing the widening term for majority carriers can be cumbersome, because re-writing the semiconductor charge density in equation (2.12) while including such a term results in an ill-conditioned expression near flatband.

The triangular approximation has been shown to be accurate for determining minority carrier and capacitance densities as a function of gate bias if more than one subband is considered [36]. Recently, it has been shown that the triangular assumption can result in an analytical expression for surface potential while considering many subbands [37]. The derivation contained in [37] is outlined and the important points presented since it will be used for quantum mechanical analysis of thin oxide devices in a later chapter.

Beginning from Poisson’s equation one immediately separates the contribution from minority carriers. After multiplying both sides by $d\psi/dy$ and integrating throughout the bulk the integral of minority carriers is considered separately. The average value of the electric field in the inversion layer ($E = -(Q'_s + Q'_i)/2\varepsilon_s$) is used to approximate the integral of minority carriers, that in turn allows for calculation of the surface field. The result contained below, implied by [37], could have been formulated by assuming $Q'_s = Q'_i + Q'_b$ with $Q'_b$ equal to equation (2.12) devoid of contributions from minority carriers.

$$Q'_s = Q'_i - \text{sign}(\psi_s)\sqrt{2qN_A\varepsilon_s}\left[\phi_i \left(e^{-\psi_s/\phi_i} - 1\right) + \psi_s\right]^{1/2} \quad (2.16)$$

The inversion charge density is given by a summation over all the subbands (in [37] it was suggested that index $j$ range from 0 to 3). The index, $i$, takes on labels $L$ and $H$ pointing to the lower and higher valleys. The quantities $m_{dL}$ and $m_{dH}$ are the density of states effective masses per valley, valued at 0.190$m_0$ in the lower valley (two-fold degeneracy) and 0.417$m_0$ in
the higher valley (four-fold degeneracy). The quantities $m_{zL}$ and $m_{zH}$ are the normal masses, valued at $0.916m_0$ in the lower valley and $0.190m_0$ in the higher valley [20]. In the proceeding equation, $E_{Fn}$ is the quasi-Fermi level, and $E_{ij}$ are energy levels corresponding to the different subbands.

$$Q'_i = -q \frac{kT}{\pi \hbar^2} \sum_i g_i m_{di} \sum_j \ln \left( 1 + e^{(E_{Fn} - E_{ij})/kT} \right)$$  \hspace{1cm} (2.17)$$

The energy levels given by $E_{ij}$ are eigenvalues resulting from the triangular well solution of Schrodinger’s equation. They are referenced to the location of the conduction band edge at the surface, $E_{c}^0$, since in the solution of Schrodinger’s equation the bottom of the well is assumed to be at zero energy.

$$E_{ij} = E_{c}^0 + \left( \frac{\hbar^2}{2m_{zL}} \right)^{1/3} \left[ \frac{3\pi q}{2} \left( j + 3/4 \right) \right]^{2/3}$$  \hspace{1cm} (2.18)$$

The average value of the electric field in the quantum well, $E$, is technically in terms of $Q'_i$, but replacing it with the surface field multiplied by a coefficient allows for straightforward numerical search for surface potential. $\eta$ is a coefficient found by [37] to be approximately 1.1. The straightforward numerical search is made possible by replacing $-Q'_s$ with $Q'_g + Q'_{ox}$ and using Gauss’s Law, which can be obtained from the relations supporting the SPE.

$$E \approx \frac{E_s}{\eta}$$  \hspace{1cm} (2.19)$$

$$= \frac{C'_0 \left( V_{GB} - V_{FB} - \psi_s \right) + Q'_H - Q'_{it00}}{\eta \varepsilon_s}$$  \hspace{1cm} (2.20)$$
The difference $E_{Fn} - E_{ij}$ in equation (2.17) can be written in terms of surface potential. Knowing that $E_{Fn} = E_F - qV$ and $E_{c}^0 = E_c - q\psi_s$ gives the desired result. The difference between bulk conduction and Fermi levels, $E_c - E_F$, is equal to $kT \ln(N_c/n_i) + q\phi_F$ with $N_c$ being the effective conduction band density of states. It can be approximated as $E_g/2 + q\phi_F$.

$$E_{Fn} - E_{ij} = q\psi_s - qV - \left(\frac{\hbar^2}{2m_zL}\right)^{1/3} \left[\frac{3\pi q (j + 3/4)E}{2} \right]^{2/3} - kT \ln(N_c/n_i) - q\phi_F \quad (2.21)$$

The above equations form the core of the analytical approximation derived in [37]. In this research work interface charge density has been included in equation (2.20) and directly affects the subbands’ energy levels. This is the only modification needed to include interface states in the one-dimensional solution of Poisson’s equation while considering quantum confinement of minority carriers.

Quantum confinement results in a reduced transconductance that is most apparent for thin oxides, which can be inferred from the reduced peak gate capacitance plotted in Figure 2.4. For thick oxides this reduction becomes very small and the classical result combined with changes in mobility parameters are adequate for modeling purposes. However, for large doping with magnitude approaching $10^{18}\text{cm}^{-3}$ a shift in the threshold voltage is predicted by the QM formulation, even for thick oxides.
Figure 2.4: C-V plots for different oxide thickness values (a) 2.5nm, (b) 25nm, (c) 50nm, showing the discrepancy between the classical and QM approaches.
2.3.1 Modifications for inversion in p-channel MOS devices

The semiconductor charge density in (2.16) changes to accommodate holes as they become minority carriers in a substrate of uniform n-type doping. The quantum confinement of accumulated majority carriers is again ignored.

\[ Q'_s = Q'_i - \text{sign}(\psi_s) \sqrt{2qN_D\varepsilon_s} \left[ \phi_s \left( e^{\psi_s/\phi_s} - 1 \right) - \psi_s \right]^{1/2} \]  

(2.22)

The inversion charge density, now comprised of holes, is evaluated differently since there are three hole bands: the light hole, heavy hole, and split-off bands. The index \( i \) takes on values \( L, H, \) and \( S \) corresponding to the three hole bands. Index \( j \) is the “rung on the energy ladder” where higher index values contribute a smaller portion to the total inversion charge density. These comments are made based on information contained in [38], which also contains the quantization and density of states effective masses for the (100) silicon surface. The total charge density is taken to be unchanged relative to the nMOS case except for the sign and argument in the exponential. Also, the split-off band has a constant energy difference between the two other bands, which must be accounted for when calculating charge [38].

\[ Q'_i = q \frac{kT}{\pi \hbar^2} \sum_i g_im_{di} \sum_j \ln \left( 1 + e^{(E_{ij} - E_{Fp})/kT} \right) \]  

(2.23)

Now, the energy levels \( E_{ij} \) are some energy below the surface valence band. The average value of the electric field is the same except the value \( \eta \) is an unknown and should be calibrated relative to data obtained by self-consistent solution of the coupled Schrodinger and Poisson equations. The triangular well approximation results in a similar solution as for the nMOS
case in (2.18).

\[
E_{ij} = E_v^0 - \left( \frac{\hbar^2}{2m^* L} \right)^{1/3} \left[ \frac{3\pi q}{2} (j + 3/4) E \right]^{2/3}
\]  

(2.24)

The difference between an energy level \(E_{ij}\) and the quasi-Fermi level also changes slightly. Knowing that \(E_{Fp} = E_F - qV\) and \(E_v^0 = E_v - q\psi_s\) gives the desired result. The difference between bulk valence and Fermi levels, \(E_v - E_F\), is equal to \(-kT \ln(N_v/n_i) + q\phi_F\) with \(N_v\) being the effective valence band density of states. It can be approximated as \(E_g/2 + q\phi_F\).

\[
E_{ij} - E_{Fp} = -q\psi_s + qV - \left( \frac{\hbar^2}{2m^* L} \right)^{1/3} \left[ \frac{3\pi q}{2} (j + 3/4) E \right]^{2/3} - kT \ln(N_v/n_i) + q\phi_F
\]  

(2.25)

### 2.3.2 Summary of useful charge density expressions

Two different semiconductor charge density results have now been presented: equations (2.12) and (2.16). The former is useful for thick oxides and lower doping where quantum mechanical considerations are of no concern. The latter considers multiple subbands and is useful for thin oxides but as shown does not handle the accumulation region. The contribution here in the present work is to leverage interface states into the surface potential equations resulting from (2.12) and (2.16). In the chapter covering compact model development only the non-quantum mechanical expression (2.12) is considered. The drain current expressions developed are still useful for thin oxides so long as the SPE stemming from (2.16) is used to compute surface potential.
2.4 Interface trapped charge

In any semiconductor device involving an insulator there are undesirable charges introduced that shift the flatband voltage, cause hysteresis of measured data, and/or modify the induced channel charge. These all manifest as a shift in the ideal flatband voltage and with the exception of surface traps do not contribute any capacitance. Insulator-related charges are categorized as: fixed, trapped, interface trap, and mobile [6]. Mobile charge produces CV hysteresis and its variation with applied voltage will be ignored.

For the purposes of modeling, a detailed understanding of fixed/trapped charge is not necessary since all appear as a shift in the flatband voltage and can be effectively lumped into the term \( Q'_o/C'_ox \). In contrast, of great interest is understanding interface traps since these charge/discharge with varying bias and can significantly affect device performance. In a simple capacitor, interface traps shift the flatband voltage and distort capacitance-voltage measurements depending on the frequency. An example diagram of interface trap occupation is shown in Figure 2.5, re-drawn from [39].

![Figure 2.5: Filled/empty states and band bending.](image-url)
2.4.1 Trap occupation

Interface states are continually distributed throughout the bandgap and are described with density $D_{it}(E)$ and probability of occupation $F(E)$. The probability of occupation varies with band bending and quasi Fermi level but is much different between the two cases of thermal equilibrium and steady-state. For thermal equilibrium the MOS device’s trap occupation function approaches a simplified form, labeled $F_o(E)$. The equilibrium $F_o(E)$ was used by Nicollian and Brews to develop MOS capacitor admittance equations for interface state analysis [6].

However, a MOSFET operating under steady-state conditions can potentially have a quasi Fermi level quite different from the bulk Fermi level owing to carrier injection from the source. The inclusion of the quasi Fermi level precludes the use of $F_o(E)$. Instead, the steady-state occupation function $F(E)$ is needed for interface trapped charge calculations unless the model is declared in advance to have limitations. The equilibrium occupation function is only valid if the amount of imref splitting is kept small relative to the surface potential. P. Muls used this limitation to investigate trap-affected MOSFET behavior in weak inversion [7].

The steady-state trap occupation probability can be given following the Shockley/Read [40] and Hall [41] formulation. The occupation function at the surface can be found if the equations are modified slightly to incorporate band-bending [27]. The usual means of writing carrier capture/emission rates involves constants of proportionality, filled/empty trap densities, and carrier concentration. The carrier concentrations are typically written in terms of Maxwell-Boltzmann statistics, and the steady-state trap occupation function reflects this by being composed of exponentials [7].

$$F(E) = \left(1 + \frac{e^{(E_i - E_F - q\psi_s)/kT} + e^{(E_i - E_i)/kT}}{e^{(E_F - E_i + q\psi_s - qV)/kT} + e^{(E_i - E_i)/kT}}\right)^{-1}$$  \hspace{1cm} (2.26)
Figure 2.6: Difference in occupation functions where \( F \) is the exact steady-state version and \( F_{FD} \) is the simplified quasi Fermi-Dirac version. The exact version indicates that \( V \) can only push the falling edge leftward until around midgap, at which point the simplified version becomes incorrect.

The above reduces only for small \( V \) to become a quasi Fermi-Dirac distribution [7].

\[
F_{FD}(E) = \left(1 + e^{(E-E_F-\psi_s+qV)/kT}\right)^{-1}
\]  

(2.27)

For small \( V \) there is negligible difference between (2.26) and (2.27). The quasi-Fermi-Dirac distribution has a falling edge that is increasingly pushed left with increasing \( V \). In contrast, while (2.26) has falling edge that is also pushed left, the degree to which it is affected saturates with large \( V \). Figures 2.6a-b demonstrates these properties. The two equations are admittedly interchangeable above weak inversion so long as \( V_{DS} \) is kept below the point of velocity saturation (onset of channel length modulation).
2.4.2 Interface trapped charge density

The total interface trapped charge involves an integral throughout the bandgap of the density function \( D_{it}(E) \) multiplied with \( F(E) \) [6]. In the derivation of (2.26) it was assumed that the traps were acceptor-like. If they had been assumed donor-like, the occupation function would simply be \( 1 - F(E) \), where \( F(E) \) is from (2.26).

\[
Q'_{it,d} = q \int_{E_v}^{E_c} [1 - F(E)] D_{it,d}(E) dE
\]
\[
Q'_{it,a} = -q \int_{E_v}^{E_c} F(E) D_{it,a}(E) dE
\]

(2.28)

(2.29)

Distinguishing between the acceptor-like and donor-like densities is deemed unimportant in this work. This is because an important observation can be made if \( Q'_{it} \) is described as the sum of \( Q'_{it,d} \) and \( Q'_{it,a} \).

\[
Q'_i(\psi_s) = q \int_{E_v}^{E_c} D_{it,d} dE - q \int_{E_v}^{E_c} F(E)(D_{it,d} + D_{it,a}) dE
\]

(2.30)

The first integral is actually bias-independent, and only shows up as a shift in the flatband voltage. The sum \( D_{it,d} + D_{it,a} \) is determined from measurements and is labeled \( D_{it} \). Therefore, from a modeling perspective, all traps can be treated as acceptor-like and the flatband voltage left as a measured quantity.

The inclusion of interface states in models often involves a simplification for the integral of interface trapped charge. One can choose to treat the trap occupation function as a step function, which takes \( F(E) \) out of the trap charge integral and modifies the upper limit of integration. This is commonly done because the integral of \( D_{it}(E) \cdot F(E) \) is often non-integrable.
\[ Q'_\text{it} \approx -q \int_{E_v}^{\psi_s + E_F - V} D_{it}(E) dE \] (2.31)

One immediate limitation of (2.31) is that \( F(E) \) temperature dependence is lost. Also, high levels of \( D_{it} \) (roughly \( > 10^{12} \text{ cm}^{-2}\text{eV}^{-1} \)) are not modeled well by this approximation unless the upper limit of integration is kept from exceeding \( E_c \). Ignoring this boundary will result in an incorrect surface potential in strong inversion, leading to an incorrect inversion charge density, leading to an incorrect drain current. Low levels of \( D_{it} \) can make use of this simplification since when the simplification is in error semiconductor charge density is dominant and interface trapping is not important. Also, (2.31) is incorrect for determining surface potential below weak inversion whenever imref splitting is included in the SPE because of the differences between (2.26) and (2.27). Consequently, equation (2.31) cannot be used for accuracy and more precise approximations to (2.30) will be generated.

Some comments regarding \( D_{it}(E) \) must be made first. It has been found experimentally that interface trap density can have a U-like distribution over the bandgap with peaks at the band edges. A function that empirically describes this behavior uses two exponentials which decay towards a midgap density [42]. The rate of decay is controlled by the bandtail parameter, denoted as \( b \).

\[ D_{it}(E) = D_v e^{-(E-E_v)/b_v} + D_i + D_c e^{(E-E_c)/b_c} \] (2.32)

Non-numerical trap-affected drain current models usually focus on the \( D_i \) term, which is constant and can be taken outside the \( Q'_{it} \) integral. However, bandtail states can introduce signifi-
cant effects in plots of drain current, the exact details of which are left for a later section.

In regards to drain current only $D_i$ and the minority carrier bandtail are important. As an example, an n-channel MOSFET at the onset of weak inversion has $\psi_s = \phi_F$ and roughly all states below midgap are occupied. Assuming the $D_v e^{-(E-E_v)/b}$ term has dropped below $D_i$ then contributions from the valence bandtail have saturated. Therefore, for an n-channel MOSFET only two terms from (2.32) are needed to capture the effects of interface states on drain current. The ignored term simply shifts the flatband voltage. For capacitance simulation from accumulation through strong inversion all terms would be needed. Example plots of $D_{it}$ and the resulting $Q'_{it}$ are shown in Figures 2.7 and 2.8.

\[
D_{it}(E) = D_i + D_v e^{(E-E_v)/b} \tag{2.33}
\]

Interface states stretch out the surface potential gate voltage relationship. This is the well-known voltage stretchout referred to in describing capacitance-voltage curves. Stretch-out also occurs in the current-voltage domain. All cases of stretch-out can be traced back to the SPE which relates system charges with gate voltage. Example plots are shown in Figure 2.9.
Figure 2.7: Effect of bandtail parameter $b$ on interface state charge density. Without the $D_c$ term $Q_{it}'$ would increase linearly with respect to $\psi_s$ and then saturate as $E_c$ is bent down towards $E_F$. $V = 0$.

Figure 2.8: Effect of different densities on interface state charge density for $V = 0$. 

31
Figure 2.9: Effect of different densities on the surface potential equation for $t_{ox} = 30$nm. In (b), solid line: $D_i = 2 \times 10^{12}$, $D_c = 5 \times 10^{13}$; dashed line: $D_i = 10^{12}$, $D_c = 10^{13}$; dot-dashed line: $D_i = 5 \times 10^{11}$, $D_c = 10^{12}$. 

32
2.4.3 Approximating trapped charge density

Because of equation (2.32) the interface trapped charge density becomes split into three components. None are integrable with the steady-state $F(E)$ but approximations are possible. The flatband-contributing offset in (2.30) is ignored since it has no bias dependence.

$$Q'_t = \text{VAL} + \text{MID} + \text{CON} \quad (2.34)$$

$$\text{VAL} = -qD_v \int_{E_v}^{E_c} e^{-(E-E_v)/b} F(E) dE \quad (2.35)$$

$$\text{MID} = -qD_i \int_{E_v}^{E_c} F(E) dE \quad (2.36)$$

$$\text{CON} = -qD_c \int_{E_v}^{E_c} e^{(E-E_v)/b} F(E) dE \quad (2.37)$$

The midgap density integral can be accurately approximated by observing graphically its derivative with respect to surface potential. The proposed solution is to use a summation of Fermi functions.

$$\frac{d\text{MID}}{d\psi_s} \approx -qD_i \left( \frac{1}{1 + e^{-(\psi_s-(E_i-E_F))/kT}} - \frac{1}{1 + e^{-(\psi_s-(E_i-E_F))/kT}} - \frac{1}{1 + e^{-(\psi_s-(E_c-E_F))/kT}} + \frac{1}{1 + e^{-(\psi_s-(E_c-E_F+V))/kT}} \right) \quad (2.38)$$

Equation (2.38) provides a useful approximation for an MOS capacitor where $V$ is constant and zero. Its indefinite integral gives total trapped charge plus an offset. The offset should be set equal to $qE_gD_i$ since this is the maximum possible trapped charge available.
\begin{align*}
  \text{MID} \approx -qD_iE_g + qD_iV_T \ln \left[ \frac{1 + e^{-(\psi_s - E_i + E_F)/kT}}{1 + e^{-(\psi_s - E_c + E_F - V)/kT}} \right] \left( 1 + e^{-(\psi_s - E_F)/kT} \right) \left( 1 + e^{-(\psi_s - E_i + E_F - V)/kT} \right) 
\end{align*}

The quality of MID is imperfect away from the $V = 0$ case. When the surface potential equals $\phi_F$ or $\phi_F + V$ the integral of the steady-state occupation function undergoes a rapid transition that is not captured perfectly by the MID approximation. Relative error peaks around half a percent (Figure 2.10), which is undesirable. However, it is tolerable because inversion charge density is very small around these two points ($\phi_F + V$ being onset of weak inversion). This peak value is independent of $D_i$ and $V$ but it is a function of temperature.

Contributions from VAL and CON are only important when they exceed MID. In the region where VAL, CON $\geq$ MID the occupation function can usually be replaced with the quasi Fermi-
Dirac version. At large values of \( V \) where the quasi Fermi-Dirac occupation function is in error, the contribution from \( \text{CON} \) is usually negligible relative to that of \( \text{MID} \) assuming \( D_i \) is only around a hundredth of \( D_c \). However, when \( b \) becomes large (roughly 0.1 eV and above) the quasi-Fermi-Dirac occupation function can give significant error when \( V > 0 \) and therefore cannot be used.

Any approximation to \( \text{CON} \) must be precise because inversion charge density is very sensitive to small changes in surface potential. For this reason the integral associated with \( \text{CON} \) is approximated by discrete integration. Applying the trapezoidal rule requires the integrand be generated for many datapoints, typically around 1000 for relative error below 0.01\%. Such a large number of points slows down the numerical search for \( \psi_s \) in the SPE increasing model computation time. A solution is to successively apply integration by parts.

\[
\text{CON} = \sigma + \sum_{k=0}^{N} (-1)^k b^{k+1} e^{(E-E_c)/b} F^{(k)} \bigg|_{E_v}^{E_c} \\
\sigma = (-b)^{N+1} \int_{E_v}^{E_c} e^{(E-E_c)/b} F^{(N+1)} dE
\]  

(2.40)  

(2.41)

The superscript attached to \( F \) in parenthesis indicates the derivative of order \( k \). \( \sigma \) is the remaining integral after integration by parts is applied \( N \) times. It is dominant when \( b > V_T \) and the Fermi level is below the conduction band edge. When \( b < V_T \) it is small and the summation converges more rapidly to the actual integral.

The proposed integration strategy is to evaluate the summation from \( k = 0 \) to \( N = 2 \) and apply Simpson’s rule over a reduced number of points to \( \sigma \). This gives peak relative error typically near a hundredth of a percent at room temperature for values of \( b \) below 0.1 eV. Simpson’s rule divides the range \([E_v, E_c]\) into an even number of subintervals and creates second order polynomials from groups of three datapoints. The resulting integration is more accurate
than the trapezoidal rule for $\sigma$. The reason that Simpson’s rule is not just applied to the integrand of CON is because the integrand’s behavior versus $E$ is significantly different depending on whether $b$ is greater or less than $kT$. The summation and $\sigma$ take turns being the dominant component of CON depending on $b$ and bias.

Error in the approximation is quantified in Figures 2.11 and 2.12 for $\sigma$ integrated over 100 points. “Exact” results are for $Q_{it}$ discretely integrated over $10^5$ datapoints. This error propagates into the SPE and slightly modifies the $V_{GB}$ axis. Since the peak error is very low there is negligible difference in predicted inversion charge density as a function of gate-bulk voltage. To visualize this the SPE can be evaluated for a range of surface potential values from midgap through strong inversion. The delta in predicted gate voltage is generally quite small for $V = 0$ and $b \leq 0.1\text{eV}$ as shown in Figure 2.13.
Figure 2.12: Magnitude of error associated with the summation and $\sigma$ integration approach for $V = 0.5$. As the ratio $D_i/D_c$ is increased the error will decrease.

Figure 2.13: Delta in predicted gate-bulk voltage by evaluating SPE at $t_{ox} = 30\text{nm}$. Increasing $V$ shifts the curves toward the right and highlights error caused by MID.
2.4.4 Expanding $Q'_{it}$

Exact analytical computation of drain current is impossible when the effects of interface states are included. Nth-order expansions are invariably required meaning that the first through Nth-order derivative of $Q_{it}$ must be known. In this work second order expansions of $V$ must be computed, thus the first and second derivatives of MID and CON are considered.

The first derivative of $Q'_{it}$ is the interface state capacitance as $f \rightarrow 0$ Hz. It is commonly approximated away from the band edges (for a MOS capacitor).

$$C_{it} = -\frac{dQ'_{it}}{d\psi_s}$$

$$\approx D_{it}(\psi_s + E_F).$$

However, the above result is not very useful for all biases since $Q'_{it}$ saturates near the band edges. More importantly, high levels of $D_{it}$ obviate its use. If the magnitude of $D_{it}$ is small then the region where the above approximation is invalid is also the region where $C_{it}$ is dwarfed by semiconductor capacitance. So, the fact that it is invalid is unimportant for low $D_{it}$. However, for high densities of interface states (particularly near the band edge) none of this is true. For this reason more exact expressions are utilized.

The derivative of $Q'_{it}$ in a MOSFET is complicated by imref splitting. This is in contrast to a MOS capacitor where $V$ is zero and constant. When a current flows in a MOSFET both $V$ and $\psi_s$ vary laterally from source to drain because of applied bias $V_{DS}$. Therefore the derivative of interface trap charge should be computed while accounting for $dV/d\psi_s$. For MID the process is straight-forward and differs from equation (2.38).
\[
\frac{d\text{MID}}{d\psi_s} \approx -qD_i \left( \theta_{m1} - \theta_{m2} \frac{dV}{d\psi_s} \right) 
\]  \hspace{1cm} (2.44)

\[
\theta_{m1} = \frac{1}{1 + e^{(E_v-q\psi_s-E_F)/kT}} - \frac{1}{1 + e^{(E_i-q\psi_s-E_F)/kT}} + \frac{1}{1 + e^{(E_i-q\psi_s-E_F+qV)/kT}} - \frac{1}{1 + e^{(E_c-q\psi_s-E_F+qV)/kT}} 
\]  \hspace{1cm} (2.45)

\[
\theta_{m2} = \frac{1}{1 + e^{(E_i-q\psi_s-E_F+qV)/kT}} - \frac{1}{1 + e^{(E_c-q\psi_s-E_F+qV)/kT}} 
\]  \hspace{1cm} (2.46)

Derivatives of CON are evaluated by distributing \( d/d\psi_s \) into the integral. Integration by parts is not applied because \( dF/d\psi_s \cdot e^{(E-E_c)/b} \) is more accurately discretely integrated over a small number of points than is \( F \cdot e^{(E-E_c)/b} \).

\[
\frac{d\text{CON}}{d\psi_s} = -q \frac{D_c}{\Phi_h} \left( \theta_{c1} - \theta_{c2} \frac{dV}{d\psi_s} \right) 
\]  \hspace{1cm} (2.47)

\[
\theta_{c1} = \int_{E_v}^{E_c} e^{(E-E_c)/b} \left( e^{(E_F-E_i+q\psi_s-qV)/kT} \frac{B-A}{B^2} + e^{(E_i-E_F-q\psi_s)/kT} \frac{A}{B^2} \right) dE 
\]  \hspace{1cm} (2.48)

\[
\theta_{c2} = \int_{E_v}^{E_c} e^{(E-E_c)/b} e^{(E_F-E_i+q\psi_s-qV)/kT} \frac{B-A}{B^2} dE 
\]  \hspace{1cm} (2.49)

Functions \( A \) and \( B \) are the numerator and denominator of the steady-state occupation function.

\[
A = e^{(E_F-E_i+q\psi_s-qV)/kT} + e^{(E_i-E)/kT} 
\]  \hspace{1cm} (2.50)

\[
B = A + e^{(E_i-E_F-q\psi_s)/kT} + e^{(E-E_i)/kT} 
\]  \hspace{1cm} (2.51)

The interface trap capacitance at very low frequency is plotted in Figure 2.14 for three values of \( b \). A unique peak is reached for every value of \( b \) even though \( D_c \) is constant for all
Figure 2.14: First derivative of $Q'_t$ for $V = 0$ (MOS capacitor case). In (b) each solid exponential line is a plot of $D_i + D_c e^{(\psi_s + E_F - E_c)/b}$ to investigate the common approximation $C_t/q \approx D_t$. The x-axis can be thought of as $E - E_v$ similar to Figure 2.7.

three curves. In Figure 2.14b the validity of $C_t/q \approx D_t$ is examined. The approximation works well for large values of $b$ but fails when $b$ is near or below $\phi_t$.

The reason the approximation fails for $b < \phi_t$ when $-dQ'_t/d\psi_s$ rises toward its peak can be explained by considering the integrand of CON. If $V = 0$ then a first order expansion about $E = E_c$ gives a good approximation to the integrand, particularly for $b < \phi_t$.

\[
e^{(E-E_c)/b} \cdot F_{FD} \approx e^{(E-E_c)/b} \left[ \frac{1}{1 + e^{(E_c-E_F-q\psi_s)/kT}} - \frac{1}{kT} \cdot \frac{e^{(E_c-E_F-q\psi_s)/kT}}{\left(1 + e^{(E_c-E_F-q\psi_s)/kT}\right)^2} \right] (E-E_c) \]

The above, when integrated to give $Q'_t$, gives a result that is negligibly dependent upon $b$. In contrast, when $q\psi_s + E_F$ is far from the band edge then the occupation function is well
approximated as a simple exponential.

\[ e^{(E-E_c)/b} \cdot F_{FD} \approx e^{(E-E_c)/b} e^{-(E-F_E-q\psi_s)/kT} \quad (2.53) \]

The above is easily integrated into an exponential with \( b \) influencing the slope on a log scale of \( Q'_H \) versus surface potential. In between values of \( q\psi_s + E_F \) far from \( E_c \) and near \( E_c \) the behavior of the integrand is not well approximated by simple expansions. Therefore, a simple description of \( -dQ'_H/d\psi_s \) and its relation to \( D_n \) is not possible near the band edges for small values of \( b \).
Chapter 3

Drain current

A flow of carriers from drain to source caused by an applied drain-source bias is associated with a gradient in the quasi-Fermi level as a function of position in the channel. Departing from the source, minority carriers flow towards the drain over a distance, $L$, in what can be assumed to be a current density that is constant across the channel width, $W$. The voltage applied to the gate electrode controls the current passing through the device and is related to the system of internal charges by the surface potential equation. The expression for the laminar flow of current in MOSFET compact models is usually taken to be,

$$I_{ds} = -\mu W Q'_i \frac{d\psi_s}{dx} + \mu W V_T \frac{dQ'_i}{dx},$$

(3.1)

which is a manipulation of the expression involving the gradient of the position dependent portion of the quasi Fermi level [28].
\[
I_{ds} = -\mu W Q'_i \frac{dV}{dx} \tag{3.2}
\]

Equation (3.1) is referred to as “drift-diffusion” and becomes an approximation of equation (3.2) in a charge sheet model [16, 43]. “Exact” one-dimensional solution of (3.2) is performed numerically and is commonly referred to as the Pao-Sah double integral. It being computationally slow to evaluate, the double integral is primarily useful for gauging model error. Equation (3.2) can become a double integral depending on how \( Q'_i \) is evaluated. Classically, it is found by integration of the minority carrier charge density from the surface to deep in the bulk. Knowing that \( d\psi_s/dy = -E \), a change of variables can be performed where \( E \) is known analytically from one of the steps used to establish (2.12). For an n-channel FET, the minority carriers are electrons.

\[
Q'_i = -qN_A \int_0^{\psi_s} \frac{e^{(\psi - 2\phi_F - V)/\phi}}{E(\psi)} d\psi \tag{3.3}
\]

Therefore, the exact expression for current that the final compact model can be compared against is the following double integral [16]. The influence of interface states is not explicit but is contained within the surface potential equation that relates \( V, \psi_s, \) and \( V_{GB} \).

\[
I_{ds} = qN_A \frac{W}{L} \int_{V_{SB}}^{V_{DB}} \int_0^{\psi_s} \frac{e^{(\psi - 2\phi_F - V)/\phi}}{E(\psi)} d\psi dV \tag{3.4}
\]

Since the above expression can only be evaluated numerically it is useful to resort to the charge sheet approximation, which treats the inversion layer as an infinitesimally thin region.
over which $\psi_s$ drops a negligible amount. This allows for the separation of $Q'_s$ into inversion and depletion components. Doing so generally allows for analytical integration (at least when ignoring interface states) of $Q'_i$ and therefore an analytical expression for drain current. It is the supporting assumption for the classic “charge sheet model” [44] and is an integral part of surface potential models. The depletion, or “bulk,” charge is found in the same manner as $Q'_s$ is derived, but with a charge density, $\rho$, devoid of minority carriers. Inversion charge density is given as the difference between $Q'_s$ and $Q'_b$.

$$Q'_b = \mp \sqrt{2qN_A \varepsilon_s} \left[ \psi_s \phi_t \left( e^{-\psi_s/\phi_t} - 1 \right) \right]^{1/2}$$ (3.5)

If the carriers are not treated classically then the QM expression for inversion charge density in Chapter 2 can be used. The drain current expressions developed below are understood to still be valid provided surface potential and $Q'_i$ are calculated appropriately.

### 3.1 Charge Sheet Approximation details

If $Q'_s$ is given as the sum of $Q'_i$ and $Q'_b$, then $Q'_i$ can be evaluated in three ways. The first method is simply by computing the difference $Q'_s - Q'_b$. A more common approach is by rearranging the SPE.

$$Q'_i = C'_{ox} \left( V_{FB} + \psi_s - V_{GB} - \frac{Q'_b + Q'_it - Q'_i00}{C'_{ox}} \right)$$ (3.6)

Equation (3.6) is normally useful for Taylor expansions of $Q'_i$ when $D_{it} = 0$ but even then is not useful for computing $Q'_i$ at low gate bias. At low gate bias, deep in weak inversion,
it is difficult to be numerically precise in the calculation of $Q'_i$ with (3.6) since the quantity in parenthesis invariably involves finding the difference of two very similar values. Precise calculation of inversion charge density is necessary to calculate low current levels. By multiplying $Q'_s - Q'_b$ by a fraction equal to unity with numerator/denominator equal to $Q'_s + Q'_b$, the following expression is obtained [16].

\[
Q'_i = 2q\varepsilon_s N_A \phi_t \frac{e^{-(V+2\phi_t)/\phi_t} \left(e^{\psi_s/\phi_t} - 1\right)}{Q'_s + Q'_b}
\] (3.7)

Since inversion charge density varies over many orders of magnitude from subthreshold to strong inversion, equation (3.7) is preferred.

Of course, assuming that the depletion region experiences the full $\psi_s$ drop is not totally correct. What matters is to what degree the assumption is incorrect. The exact computation of $Q'_i$ involves an integration as written in (3.3). A similar integral can be written for the bulk charge where $e^{(\psi - 2\phi_t - V)/\phi_t}$ is replaced by $1 - e^{-\psi/\phi_t}$. The vertical electric field, $E$, is simply $-Q_s/\varepsilon$ evaluated at an arbitrary $\psi$ instead of $\psi_s$.

A comparison between the numerically integrated result (labeled “exact”) and the charge sheet approximation is given in Figure 3.1b. Clearly, the approximation is valid by producing minimal error in the range of surface potentials where current is detectable (weak through strong inversion).
Figure 3.1: Surface potential swept from $\phi_F$ to $3\phi_F$ (an extreme value but chosen to show the decrease in error). In (a), the largest curve corresponds to the greatest doping and vice-versa.
As shown in Figures 3.1a-b some doping and bias-dependent error is unavoidably included by making the charge sheet approximation. More error is included by using the drift-diffusion equation for $I_{ds}$ contained in (3.1). In the context of a charge sheet model the difference between (3.1) and (3.2) also contributes some error, since there is some amount of approximation involved in the statement that $dQ_i'/d\psi_s = Q_i' / \phi_t [1 - dV/d\psi_s]$. Adding in interface states to the mix further compounds the situation and two plots are created to gauge the usefulness of (3.1) in Figure 3.2.

The error is calculated by comparing the drift-diffusion equation (3.1) with the charge sheet approximation against (3.4) for a fictional long-channel device with unity $W/L$. A fixed and small $V_{DS} = 10$ mV is chosen for demonstration and the gate voltage swept. The mobility is set to a constant value of 200 cm$^2$/Vs and the comparison is started from the point where $I_{ds}$ is close to a fA and stopped near a uA. Interestingly, the peak error for the trap-affected data is lower than the trap-free case. The error changes with different values for the components of the $D_{t}$ expression.
Figure 3.2: Percent error of the drift-diffusion charge sheet model for $I_{ds}$ relative to the numerically evaluated Pao-Sah expression. The x-axis is different in (b) because of voltage stretchout by $D_{it}$. 

(a) $D_{it}=0$

(b) $D_{i}=5 \times 10^{11}, D_{c}=3 \times 10^{13}, b=0.1$
3.2 Behavior of trap-affected $Q'_i$

An important distinction must be made when discussing the derivation of a trap-affected compact model, specifically that it is much more involved than a trap-free model. In the absence of interface states it is trivial to analytically integrate equation (3.6) with respect to surface potential (by ignoring the exponential term in $Q'_b$). This is because (3.6) written without $Q'_it$ is completely devoid of imref splitting, $V$. Analytical integration is then possible or one can resort to simple expansions with respect to $\psi_s$. However, when one is interested in significant $D_{it}$ levels the presence of $Q'_i$ brings $V$ back into (3.6) since the occupation function used to calculate $Q'_it$ is a function of $V$.

Without interface states, the behavior of inversion charge density versus surface potential is approximately linear. This feature opens the door to very compact expressions of drain current based on the Symmetric Linearization Method (SLM) [45] utilized in several successful models [16, 46, 15]. However, high levels of $D_{it}$ cause deviations from linearity. The presence of $Q'_it$ produces a $Q'_i, \psi_s$ relationship that is not nearly as consistent. Depending on the gate, drain, and source biases the relationship can approach an exponential, while at other bias conditions the relationship can become more linear. Between these two extremes $Q'_i$ vs $\psi_s$ is neither purely exponential nor linear, a difficult-to-describe relationship that tends to occur in the sub-threshold to super-threshold transition on a plot of $I_{ds}$ vs $V_{GS}$. The boundary between the two extremes occurs gradually and is difficult to define because it depends on oxide thickness, trap parameters, and terminal voltages.

Plots accompanying the above discussion are shown in Figures 3.3a-d. To generate each figure the source-side potential ($\psi_{s0}$) was forced to equal values near twice the Fermi level. Then, the SPE was computed to determine the associated $V_{GB}$. With $V_{GB}$ known, the drain-end surface potential ($\psi_{sL}$) was computed numerically from the SPE for $V_{DS} = 0.5V$. Letting $\psi_s$
vary from $\psi_{s0}$ to $\psi_{sL}$ the channel potential $V$ was computed numerically from the SPE at every value of $\psi_s$. Finally, $Q'_i$ was computed from equation (3.7). The axes are scaled so that every line has the same $x$ and $y$ range, done because inversion charge density varies exponentially. The trap-affected figures have curves that can be thought to have localized linear regions that slowly transition to an exponential region. Non-linearity is overcome as the drain end inversion charge density nears the drain end interface state charge density.
Figure 3.3: Legend indicates value of $\psi_s$ (source-side surface potential). (a) demonstration of linearity when traps are absent. (b)-(d) traps cause non-linearity until a large enough gate bias is applied (strong inversion triode region). $V_{DS} = 0.5V$, $t_{ox} = 30nm$, $N_A = 10^{17} \text{cm}^{-3}$. $Q_{i0,L} = Q_i(\psi_{s0,L}, V_{SB, DB})$. 

$d_i = 0$, $D_c = 0$

$D_i = 10^{11}$, $D_c = 10^{13}$, $b = 0.02$

$D_i = 10^{11}$, $D_c = 10^{13}$, $b = 0.05$

$D_i = 10^{11}$, $D_c = 10^{13}$, $b = 0.1$
3.3 Effective Mobility

Solving (3.1) involves multiplying both sides by $dx$ and integrating. Assuming steady-state DC, drain current is constant throughout channel and one finds that inversion charge density must be integrated with respect to surface potential. The quantities $\psi_{sL}, \psi_{s0}$ are the drain and source-side surface potentials calculated numerically from equation (2.14) with $V = V_{DB}, V_{SB}$.

$$I_{ds} = \frac{W}{L} \left[ - \int_{\psi_{s0}}^{\psi_{sL}} \mu Q'_i d\psi_s + V_T \int_{Q_{i0}}^{Q_{iL}} \mu dQ'_i \right]$$  \hspace{1cm} (3.8)

The drift-diffusion equation was written provided the MOSFET is operating under conditions that satisfy the Gradual Channel Approximation (horizontal field very small relative to vertical field). This allowed for the velocity-field relationship to be written as $v = \mu \cdot \frac{d\psi_s}{dx}$ as reflected in (3.1). However, mobility is a function of vertical field and can change significantly as a function of gate bias. Scattering effects in the channel region produce a mobility-field relationship that gives values much different from a bulk mobility value. In compact models the carrier dependency on effective vertical field is usually expressed in the form of,

$$\mu = \frac{\mu_0}{1 + f(E_e)}.$$  \hspace{1cm} (3.9)

The exact form of $f(E_e)$ varies and depends on the scattering mechanisms under consideration. It is more desirable to leave it as a generic function to be extracted from measurement. $E_e$ is the effective vertical field equal to the average above and below the channel [47]. The relative permittivity is that of the semiconductor. $\eta$ allows for a “universal” mobility-field relationship that predicts high-field mobility regardless of changes in electrical behavior caused
by the body effect ($V_{SB} \neq 0$).

$$E_e = \frac{1}{\varepsilon} (\eta |Q'_i| + |Q'_b|)$$ (3.10)

Since mobility is clearly a function of surface potential by way of $Q'_i$ and $Q'_b$, solution of drain current becomes more complicated because $\mu$ is inside the integrals of (3.8). It is usually impossible to include a field-dependent $\mu$ and arrive at an analytical $I_{ds}$ expression without any simplification. A notable exception is the work of J.R. Hauser [48] whose results are valid in inversion for mobility degradation by surface roughness. For a non-regional model it is instead common to replace $\mu$ with an effective and constant quantity, $\mu_e$. For such a simplification the mobility is evaluated at a convenient surface potential in the channel and removed outside each integral.

$$I_{ds} \approx \mu_e \frac{W}{L} \left[ - \int_{\psi_s}^{\psi_{L}} Q'_i d\psi_s + V_T \int_{Q_{i0}}^{Q_{iL}} dQ'_i \right]$$ (3.11)

A possibility for $f$ is that it is proportional to $E_e^2$, which is the case for scattering by surface roughness. In [48], $f(E_e) = \frac{\mu_0}{K_{sr}} E_e^2$ where $\mu_0$ is the low field mobility and $K_{sr}$ is a constant. For an appropriately sized $K_{sr}$ value, $\mu$ becomes approximately linear as a function of laterally-varying $E_e$ at any gate bias. Effectively, the product of field-dependent $\mu$ and $Q'_i$ becomes equivalent to a reduced and slightly tilted $Q'_i$. To capture the reduction and tilt, the integral of $Q'_i$ can be multiplied by the average mobilities at its endpoints, or simply its value at the channel midpoint. Validity is demonstrated in Figure 3.4.
Figure 3.4: Effect of field-dependent mobility on channel inversion charge. $r$ equals $\mu_e$ evaluated at the channel midpoint, which clearly produces a convenient alternative to integrating $\mu Q_i'$. The mobility-affected quantities were normalized by $\mu_0 = 350\text{cm}^2/\text{Vs}$; also, $K = 6(10^{14})\text{V/s}$. Gate bias chosen to operate simulated device in the triode region.

### 3.4 Calculating drain current

Figures 3.3a-d demonstrated the non-linearity of inversion charge density as a function of surface potential. Care must be taken when approximating the integral of $Q_i'$ because increasing values of $V_{DB}$ do not cause $\psi_{sL}$ to quickly saturate. This is in contrast to the trap-free case where $\psi_{sL}$ is rapidly pinned to an upper limit. With interface states the drain-end surface potential will continue to rise, effectively lengthening the x-axis in a plot of $Q_i'$ vs $\psi_s$. Therefore, any expression approximating $\int Q_i' d\psi_s$ that involves the term $\psi_{sL}$ runs the risk of producing unphysical negative differential behavior in a plot of $I_{ds}$ vs $V_{DS}$.

To avoid unintended negative differential behavior a “pinchoff” surface potential is defined as can be done for the trap-free case but modified for presence of $Q_{it}'$. Here, (3.6) is written with $Q_{it}'$ equal to its source-side value, $Q_{it0}$. The resulting equation thereby becomes independent of $V$. It can be set equal to $Q_{itL}$ and solved for the corresponding value of surface potential, $\psi_p$. 

\[ \int Q_i' d\psi_s = Q_{itL} \]
provided \( Q'_b \) is removed of its negligible exponential component.

\[
\psi_p = \left( -\sqrt{2q\varepsilon_s N_A} + \sqrt{2q\varepsilon N_A - 4C'_{ox}(c - Q_{iL} + C'_{ox}V_T\alpha)} \right) + V_f \quad (3.12)
\]

\[
c = C'_{ox}\left( V_{FB} - V_{GB} - \frac{Q_{i0} - Q_{i00}}{C'_{ox}} \right) \quad (3.13)
\]

Allowing \( \psi_p \) to act as a boundary permits the integral of \( Q'_i \) to be broken up into two terms. The first term is based on the assumption that \( Q'_i(\psi_s) \) is approximately linear on the interval \([\psi_{s0}, \psi_p]\). The second term attempts to account for any remaining non-linear contribution and can be thought of as a correction. This strategy provides a smooth transition because \( \psi_p \) naturally tends towards \( \psi_{sL} \) in the triode region, which is where linearity is observed. By basing the model on \( \psi_p \) it is possible to avoid negative differential behavior provided CORR is carefully constructed.

\[
\int_{\psi_{s0}}^{\psi_{sL}} Q'_i d\psi_s \approx \int_{\psi_{s0}}^{\psi_p} \text{LIN} \ d\psi_s + \int_{\psi_p}^{\psi_{sL}} \text{CORR} \ d\psi_s \quad (3.14)
\]

The linear region between \( \psi_{s0} \) and \( \psi_p \) is well approximated by a first order expansion around a convenient location. Applying the concept of symmetric linearization at the point \((\psi_{s0} + \psi_p)/2\) results in a simple result for the integral of LIN. The surface potential given by \((\psi_{s0} + \psi_p)/2\) can be thought of as the localized channel midpoint, \( \psi_m \). The inversion charge density at \( \psi_m \) can be calculated once the SPE is numerically solved for the corresponding value of \( V \), labeled \( V_m \). Writing an expansion centered about \( \psi_m \) greatly simplifies the resulting integral.
Outside the $[\psi_{s0}, \psi_p]$ interval the inversion charge gradually transitions to an exponential. The speed with which it transitions is determined by the density and decay contained in $D_c e^{(E-E_c)/b}$. To avoid integrating with respect to $\psi_s$ (to avoid unintended negative differential behavior) a change of variables to $dQ_i'$ can be made in the integral of CORR. To perform the change of variables equation (3.7) can be considered. Assuming the terms in the denominator do not change much with respect to $\psi_s$ the derivative of (3.7) simply involves an exponential.

\[
\frac{dQ_i'}{d\psi_s} = \frac{Q_i'}{\phi_t} \left( 1 - \frac{dV}{d\psi_s} \right) \quad (3.16)
\]

This result can alternatively be obtained by analysis of the SPE after invoking a number of practical simplifications [16]. It is the basis for equation (3.1) in a charge sheet model, which in its true form involves a single derivative equal to $dV/dx$. Performing the change of variables one finds that the integral of $Q_i'$ with respect to $\psi_s$ can alternatively be expressed as an integral involving $dV/d\psi_s$ with respect to $Q_i'$.

\[
\int_{\psi_p}^{\psi_s} \text{CORR} \, d\psi_s \approx \int_{Q_{ip}}^{Q_{sL}} \frac{\phi_t}{1 - \frac{dV}{d\psi_s}} \, dQ_i' \quad (3.17)
\]

The quantity $(1 - dV/d\psi_s)^{-1}$ can be approximated by a first order expansion with respect to $Q_i'$ at the point $Q_{ip}$ (inversion charge density at $\psi_p$). To accomplish this the first and second order derivatives of imref splitting with respect to surface potential must be quantified. Both
are found starting from the SPE.

\[
\int \text{CORR} \, d\psi_s \approx \phi_t (Q_{iL} - Q_{ip}) \left( 1 - \frac{dV}{d\psi_s} \bigg|_{\psi_p} \right) + \frac{\phi_t^2 (Q_{iL} - Q_{ip})^2 \frac{d^2V}{d\psi_s^2} \bigg|_{\psi_p}}{2Q_{ip} \left( 1 - \frac{dV}{d\psi_s} \bigg|_{\psi_p} \right)^3}
\]  \hspace{1cm} (3.18)

In the final drain current expression \( \mu_e \) is evaluated at the point \( \psi_m = (\psi_{s0} + \psi_p)/2 \). Let \( M^{-1} = 1 - \frac{dV}{d\psi_s} \bigg|_{\psi_p} \) and \( \theta = \frac{d^2V}{d\psi_s^2} \bigg|_{\psi_p} / (2Q_{ip}) \). Although the resulting \( I_{ds} \) expression appears long, it practically reduces away from the subthreshold to super-threshold transition region. For very small band-bending the last, diffusion term dominates. In strong inversion the middle terms contribute very little as \( \psi_p \to \psi_{sL} \).

\[
I_{ds} \approx \mu_e \frac{W}{L} \left[ -Q_{im}(\psi_p - \psi_{s0}) - \phi_t (Q_{iL} - Q_{ip})M - \phi_t^2 (Q_{iL} - Q_{ip})^2 \theta M^3 + \phi_t (Q_{iL} - Q_{ip}) \right]
\]  \hspace{1cm} (3.19)

Percent error is demonstrated in Figures 3.5a-c relative to discretely integrated result for three oxide thicknesses. Oxides with \( C'_{ox} \) smaller than \( qD_{it} \) (discussed in Chapters 4 and 5) are more challenging to approximate for large values of \( b \) close to 100meV due to exaggerated non-linear \( Q'_t(\psi_s) \) behavior.

The downside of equation (3.19) is that it requires four numerical solutions of the SPE to find \( \psi_{s0}, \psi_{sL}, V_m, V_p \). This is estimated to take twice as long as any other trap-free surface potential-based model, which would only require numerical search for \( \psi_{s0}, \psi_{sL} \). A faster alternative to (3.19) can be created at small \( V_{DS} \) by recognizing that \( Q'_t \) versus \( \psi_s \) is well-approximated by a second order polynomial. Consequently, Simpson’s rule can be applied over three surface potentials given by \([\psi_{s0}, (\psi_{s0} + \psi_{sL})/2, \psi_{sL}] \). This result is actually more
Figure 3.5: Percent error of expression (3.19) relative to discretely integrated (3.11). Lines without markers are for $V_{DS} = 100\text{mV}$ and lines with markers are for $V_{DS} = 1\text{V}$. Higher $V_{DS}$ increases error slightly at larger band-bending. Calculated for special case of unity $\mu_e L$ to demonstrate quality of approximation to $\int Q_i' d\psi_s$. 

58
accurate than (3.19) at low drain bias and is therefore very useful for parameter extraction. Let \( \Delta \psi = \psi_{s0} - \psi_{sL} \) and \( Q_{im} \) be inversion charge density evaluated at \( (\psi_{s0} + \psi_{sL})/2 \).

\[
\downarrow I_{ds} \approx \mu_e \frac{W}{L} \left[ \frac{\Delta \psi}{6} (Q_{i0} + 4Q_{im} + Q_{iL}) + \phi_t (Q_{iL} - Q_{i0}) \right]
\]

(3.20)

Percent error at small drain-source bias is demonstrated in Figure 3.6 relative to discretely integrated result for three oxide thicknesses calculated for the special case of unity \( \mu_e \frac{W}{L} \). The approximation works quite well and keeps relative error below 0.01 (percent error below unity). For higher values of \( D_i' \) the error increases in subthreshold because of the approximation to MID contained in equation (2.39).
Figure 3.6: Percent error of expression (3.20) relative to discretely integrated (3.11) for $V_{DS} = 100\text{mV}$. 
3.4.1 Effects of $D_{it}$ on DC drain current

Since $Q'_{it}$ assists in the balance of gate charge (along with semiconductor charge) it induces gate-voltage stretch-out, something that is readily seen in plots of MOS capacitance. This phenomena also manifests in the I-V realm as a higher subthreshold swing (reduced subthreshold slope), which is a well known behavior. However, high levels of $D_{it}$ that rapidly increase near the minority carrier band edge also elongate the subthreshold to super-threshold transition. This could be termed “soft” threshold behavior and is most obvious for high values of $D_c$.

Considering Figure 3.7 one can see the reduction in subthreshold slope caused by a midgap trap density. Above subthreshold the contribution from $D_i$ to charge density saturates and one simply observes a shift in the extrapolated threshold voltage. This is why it is somewhat useless to develop a trap-affected model to analyze above-threshold data while only considering midgap states: a trap-free model could be used provided $V_{FB}$ is a parameter.

However, as large values of $D_c$ are introduced some curvature develops. Curvature makes it difficult to perform linear extrapolation for determining threshold voltage to an extent that the concept of threshold becomes relatively meaningless. An extrapolation at high $V_{GS}$ will predict a threshold that can be a significant distance away from an observed “pinch” on a linear scale (roughly 1.5V in Figure 3.7a). For the same $D_c$ value curvature can also be introduced by decreasing the rate of decay in the U-like $D_{it}$ distribution over energy (Figure 3.8).

Since the contribution from $D_{it}$ in the SPE is divided by $C'_{ax}$ the effects of high $D_{it}$ are more apparent for large oxide thicknesses. Figure 3.9 shows changes in $I_{DS}, V_{GS}$ for the same interface state density but different oxide thickness. The soft threshold behavior is less obvious as $t_{ox}$ reduces below 10nm. Even so, it is not always possible to fit measured data from subthreshold through super-threshold without the $D_c$ parameter.

A “soft” threshold could equally well be described as $g_m$-broadening. A trap-free curve’s
transconductance is compared to several trap-affected curves in Figures 3.7c, 3.8c. Without traps the increase in $g_m$ is very abrupt, whereas with traps the increase is more gradual and delayed. This behavior highlights even moreso the inability of a trap-free model to match measured data from non-ideal MOSFETs.
Figure 3.7: Effect of increasing $D_c$ with $\mu_e W / L = 1 \text{ cm}^2 / \text{V} \cdot \text{s}$, $N_A = 10^{17}$, $b = 50 \text{ meV}$, $D_i = 5 \times 10^{11}$. The x-axis is shifted for each curve by its initial value. $V_{GS}$ for each curve computed by starting and ending at source-side surface potentials $\phi_F$ and 1.0V.
Figure 3.8: Effect of increasing $b$ with $\mu_e W L = 1 \text{ cm}^2 / \text{V} \cdot \text{s}$, $N_A = 10^{17}$, $D_c = 3(10^{13})$, $D_i = 5(10^{11})$. The x-axis is shifted for each curve by its initial value. $V_{GS}$ for each curve computed by starting and ending at source-side surface potentials $\phi_F$ and 1.0V.
Figure 3.9: Increasing values of $t_{ox}$ reduce oxide capacitance density and strengthen the impact of $D_d$. 

$V_{DS} = 100mV, D_i = 0, b = 0.05$

to $x = 50nm$
to $x = 10nm$

$D_c = 3 \times 10^{13}$

$D_c = 0$
3.4.2 Velocity saturation

The drift-diffusion equation was written with the simplifying assumption that \( v = \mu E \), implying unlimited carrier velocity. However, a semiconductor velocity-field curve is non-linear and tends towards a limiting value. Therefore, not only is carrier flow disturbed by the vertical field (enclosed in \( \mu_e \)), it is also affected by the horizontal field owing to \( V_{DS} \). III-V materials such as GaAs and GaN are unique in that their velocity-field relationships exhibit a peak, which is in contrast to Si where the velocity-field curve increases monotonically towards a saturated value. Plots for each material are shown in Figure 3.10; GaN and GaAs from [49], Si from [50].

For Si, a common way to model this behavior is by re-writing \( v = \mu E \) into an expression that smoothly connects the linear \( v = \mu E \) region with \( v = v_{\text{sat}} \). The following expression does not ensure symmetry (source/drain interchange) but is easiest to implement. \( E_c \) represents a tunable parameter.

![Figure 3.10: Velocity-field curves for three materials (references in text).](image)
\[ v = \mu \frac{d\psi_s}{dx} \left[ 1 + \frac{1}{E_c \frac{d\psi_s}{dx}} \right]^{-1} \]  

(3.21)

The smooth, two-region model is unlikely to be compatible with other materials for all bias conditions because the velocity-field curve peaks and then decreases towards a saturating value with increasing field. A model that captures the peaked nature of GaN’s velocity-field curve was proposed in [51] but can only be implemented in a numerical simulator without resorting to a regional approach [52]. To compromise, one can select the velocity-field model in equation (3.21) that is accurate until the peak [53].

The drift-diffusion equation can be easily solved with the new velocity relation in equation (3.21). Let the original expression in (3.19) be labeled \( I_{ds}' \). The quantity \( E_c \) sets the “knee” of the velocity-field curve (roughly speaking where the linear \( v = \mu E \) fails).

\[ I_{ds} = K_{Sat} I_{ds}' \]  

(3.22)

\[ K_{Sat} = \frac{1}{1 + \frac{1}{E_c T} (\psi_{sL} - \psi_{s0})} \]  

(3.23)

3.4.3 Series resistances

The sheet and contact resistances associated with the source/drain can be large enough to degrade device performance due to voltage drops that diminish the applied drain-source voltage. Degradation is only noticeable at larger values of drain current and hence at larger \( V_{DS} \). The discussion below will only apply to (3.19) and not (3.20) since at very small \( V_{DS} \) the effect of series resistances is assumed negligible. To account for parasitic series resistances the channel current will first be considered above threshold and in the linear regime. Then, the current
becomes SLM-like with \( \psi_m = (\psi_{s0} + \psi_{sL})/2 \).

\[
I_{ds} = K_{Sat} \mu_m \frac{W}{L} (mV_T - Q_{im}) \Delta \psi
\]  
\( (3.24) \)

\[
\Delta \psi = \psi_{sL} - \psi_{s0}
\]  
\( (3.25) \)

\[
m = \frac{dQ'_i}{d\psi_s} \bigg|_{\psi_m}
\]  
\( (3.26) \)

Making use of an SLM-like current is a temporary substitute, eventually permitting incorporation of series resistances in a compact way. Following the work shown in [16], the source and drain-end surface potentials are replaced by \( \psi_{s0} = \psi_{s0} + I_{ds}R_s \) and \( \psi_{sL} = \psi_{sL} - I_{ds}R_s \) (assuming \( R_s = R_d \)). Then, only \( \Delta \psi \) contains the effects of series resistance due to the inherent model symmetry. Accounting for series resistance in the SLM-like expression and gathering terms a quadratic is found.

\[
0 = \frac{2R_s}{E_c L} I_{ds}^2 - I_{ds} \left( 1 + \frac{\Delta \psi}{E_c L} + 2AR_s \right) + A\Delta \psi
\]  
\( (3.27) \)

\[
A = \mu_m \frac{W}{L} (mV_T - Q_{im})
\]  
\( (3.28) \)

To avoid any square roots \( I_{ds} \) is solved for as a continued fraction: \( x = -c/(ax + b) \rightarrow x = -c/(b - ac/(ax + b)) \) . . . yielding \( x \approx -c/(b - ac/b) \).

\[
I_{ds} = \frac{A\Delta \psi}{1 + \frac{\Delta \psi}{E_c L} + 2AR_s - \frac{2R_s A\Delta \psi}{E_c L + \Delta \psi + 2R_s E_c L}}
\]  
\( (3.29) \)

Since series resistances are typically only important in strong inversion the denominator is
applied to the original drain current expression of (3.19) labeled $I'_{ds}$.

\[
I_{ds} = K I'_{ds} \tag{3.30}
\]

\[
K = \left[ 1 + \frac{\Delta \psi}{E_c L} + 2AR_s - \frac{2R_e A \Delta \psi}{E_c L + \Delta \psi + 2AR_e E_c L} \right]^{-1} \tag{3.31}
\]

Where $\Delta \psi$ and $m$ are replaced with new values that corresponds to the region where $A$ is valid.

\[
\Delta \psi = \psi_p - \psi_{s0} \tag{3.32}
\]

\[
m = \frac{Q_{IL} - Q_{IP}}{\Delta \psi} \tag{3.33}
\]

The original drain current expression has now been modified to include a compact description of series resistance and velocity saturation. Both are capable of decreasing the expected current level depending on bias and channel length. Because of interface states all midpoint values contained in the expression for $A$ are evaluated at $(\psi_{s0} + \psi_p)/2$. If it is desired that velocity saturation be ignored then $E_c \to \infty$ and $K$’s denominator reduces to $1 + 2AR_s$.

Figure 3.11 shows the effect of large series resistance and how it can significantly decrease current levels. For low values of $R_s$ the effects of $R_s$ are minor at low $V_{DS}$. Figure 3.12 shows how series resistances modify $I_{DS}$ vs $V_{GS}$ plots. A large $R_s$ mimics the behavior of effective mobility reduced by surface roughness by causing a peak in a plot of transconductance. Once in subthreshold every curve behaves as if $R_s = 0$. Without independent test structures it is difficult to separate the effects of $\mu_e$ from $R_s$ at large fields. Therefore, from a measurement/extraction perspective, all degradation in current can be attributed to $\mu_e$. 
Figure 3.11: Characteristics of a simulated device with $t_{ox} = 10\text{nm}$, $N_A = 10^{17}$, $D_i = 0$, $D_c = 10^{13}$, $b = 0.05$, $\mu_e = \mu_0 = 150$. $V_{GS}$ from top to bottom: $[3.6, 4, 4.6]$ V.

Figure 3.12: Characteristics of a simulated device with $t_{ox} = 10\text{nm}$, $N_A = 10^{17}$, $D_i = 5(10^{11})$, $D_c = 3(10^{13})$, $b = 0.05$, $\mu_e = \mu_0 = 150$, $V_{DS} = 0.1\text{V}$.
3.4.4 Current saturation

The saturation region in a plot of $I_{ds}, V_{DS}$ corresponds to a saturated drain-end surface potential. If $V_{DB}$ is raised high enough then the drain-end inversion charge can be thought to disappear and the charge at $x = L$ becomes solely in terms of $Q'_b$, which is $V_{CB}$-independent. The drain-source bias at which this occurs is termed the saturation voltage and increasing $V_{DS}$ past this point gives rise to channel length modulation.

The inversion charge density near the drain does not disappear entirely but instead reaches a minimum. The minimum inversion charge realized at the drain is governed by the steady-state current and the saturation velocity, $v_{sat}$. As $V_{DB}$ is increased the inversion charge density decreases and the carrier velocity increases in order to maintain the steady-state current. Once $v_{sat}$ is reached the inversion charge density stops decreasing. The velocity saturated current at $x = L$ is,

$$I_{ds,sat} = -Wv_{sat}Q_{IL,sat}$$ (3.34)

The drain-bulk potential is restricted once the saturated velocity is reached so that $Q_{IL}$ does not decrease further. Further increases in $V_{DB}$ are handled by using a smoothing function which chooses the minimum between the applied $V_{DB,a}$ and the calculated saturation point [16].

$$V_{DB} = \frac{V_{DB,a}V_{DB,sat}}{\left(V_{DB,a}^m + V_{DB,sat}^m\right)^{1/m}}$$ (3.35)

A pseudo-analytical means of obtaining the saturation voltage is shown next and is partly achieved by following the exact process used to determine current. The method is not com-
pletely analytical since it relies on calculation of source-side surface potential. If equation (3.20) is set equal to (3.34) then $Q_{\text{il, sat}}$ can be solved for assuming $Q_{\text{im}} \approx (Q_{\text{il}} + Q_{\text{i0}})/2$. $\mu_e$ is estimated at its source-side value and $\Delta \psi \approx \psi_{s0} - \psi_p$ where $\psi_p$ is calculated as in equation (3.12) but with $Q_{\text{il}} \approx Q_{\text{i0}}/100$.

$$Q_{\text{il, sat}} \approx \left( \frac{-L_v \psi_{\text{sat}}}{\mu_e} - \frac{\Delta \psi}{2} - \phi_t \right)^{-1} \left( \frac{\Delta \psi}{2} Q_{\text{i0}} - \phi_t Q_{\text{i0}} \right)$$  \hspace{1cm} (3.36)

The saturated drain-end inversion charge density, $Q_{\text{il, sat}}$ can be related to a surface potential and drain-bulk potential through the SPE after making an estimate for $Q_{\text{itL}}$. If equation (3.7) is rewritten with

$$K = 2q\varepsilon N_A \phi_t e^{-2\phi_p/\phi_t} \left[ Q_{\text{il, sat}} + 2Q'_{\text{b}}(\psi_p) \right]$$

then an estimate $\Delta = \psi_{sL, \text{sat}} - V_{\text{DB, sat}}$ can be obtained from $\Delta = \phi_t \ln(Q_{\text{il, sat}}/K)$. $Q_{\text{itL}}$ is then approximated as $Q'_{\text{it}}(\psi_s = \Delta, V = 0)$.

$$\psi_{sL, \text{sat}} \approx \left[ \frac{-\gamma + \sqrt{\gamma^2 - 4c}}{2} \right]^2 + V_T \alpha$$  \hspace{1cm} (3.37)

$$V_{\text{DB, sat}} \approx \psi_{sL, \text{sat}} - \Delta$$  \hspace{1cm} (3.38)

$$c = V_{\text{FB}} - V_{\text{GB}} - \frac{Q_{\text{il, sat}} + Q'_{\text{b}}(\psi_s = \Delta, V = 0) - Q_{\text{i00}}}{C'_{\text{ox}}} + V_T$$  \hspace{1cm} (3.39)

$$\gamma = \frac{\sqrt{2q\varepsilon N_A}}{C'_{\text{ox}}}$$  \hspace{1cm} (3.40)

The given estimate for $V_{\text{DB, sat}}$ is in error for small channel lengths (Figure 3.13). In a long-channel FET velocity saturation corresponds to a $Q_{\text{il}}$ that is very small relative to $Q_{\text{i0}}$. However, in short-channel devices the current is increased because of small $L$ and the drain-end inversion charge density is comparable to $Q_{\text{i0}}$. Consequently, the modified $\psi_p$ is overestimated.
and \( Q_{L,sat} \) is significant underestimated.

![Graph](image)

Figure 3.13: Top curve to bottom: \( L = 100, 10, 1 \) um. Markers: numerical search; line: analytical expression. \( N_A = 10^{17} \text{cm}^{-3}, t_{ox} = 30 \text{ nm}, \mu_e = \mu_0 = 100. \) Threshold is near 2.7V.
3.5 Model Summary

The DC models rely on the surface potential equation (SPE) of equation (2.14) to calculate the source and drain-end surface potentials. With both known all other variables such as inversion charge density can be calculated. The inclusion of interface states in the SPE eliminated the normal relationship between inversion charge density and surface potential, which provided motivation for a new strategy of computing the integral of inversion charge density with respect to surface potential.

The presented DC current expressions are valid from subthreshold to super-threshold which makes them unique compared to other compact trap-affected drain current formulations. They capture both the reduced subthreshold slope and non-abrupt threshold caused by interface states. Of the two expressions for $I_{ds}$ one is valid at small $V_{DS}$ and the other valid at all $V_{DS}$. The former is more accurate at low drain biases.

1. For small $V_{DS} \leq 0.1$. Velocity saturation is not an issue and the denominator of $K$ reduces to $1 + 2AR_s$.

$$I_{ds} \approx \mu_e K \frac{W}{L} \left[ \frac{\Delta \psi}{6} (Q_{i0} + 4Q_{im} + Q_{iL}) + \phi_t (Q_{iL} - Q_{i0}) \right]$$

(3.41)

2. For all biases.

$$I_{ds} \approx \mu_e K \frac{W}{L} \left[ -Q_{im} (\psi_p - \psi_{s0}) - \phi_t (Q_{iL} - Q_{ip}) M - \phi_t^2 (Q_{iL} - Q_{ip})^2 \theta M^3 + \phi_t (Q_{iL} - Q_{i0}) \right]$$

(3.42)

In order to evaluate either expression the SPE must be numerically evaluated twice to find $\psi_{s0}$ and $\psi_{sL}$. Since $Q'_{iL}$ was calculated numerically (although efforts were made to reduce the number of points needed) the numerical search is slower than normal. Also, it is necessary
to numerically solve the SPE for values of $V$ at $\psi_m$ and/or $\psi_p$ (an unavoidable consequence of high $D_n$). Therefore, the presented model requires more computation time than a typical surface potential based model. However, for single-transistor analysis and parameter extraction computation time is not an issue.
Chapter 4

Mobility analysis

In order to use the trap-affected DC model in the previous chapter certain device parameters must be extracted. One of the most important is the FET channel mobility, a value that varies with gate bias. Mobility is of foremost importance since incorrect evaluation can lead to incorrect values of the parameters affecting semiconductor charge. In Chapter 3 mobility was written as a field-dependent quantity with dependence in the form of $\mu_0/(1 + f(E))$ where the function $f(E)$ and low field value $\mu_0$ are determined empirically. In this chapter the extraction of both quantities is discussed beginning with the trap-free case and then the trap-affected case.

Carrier mobility in a semiconductor is dependent on the conductivity effective mass and mean time between scattering events [39]. The value of bulk carrier mobility is different from that in the MOSFET channel because of additional bias dependent scattering events. The time between scattering events in the channel is affected by the vertical field originating from the gate, a relationship found out empirically. In a MOSFET at high vertical fields the mobility is reduced because of surface roughness, while at lower fields the mobility is a function of phonon and impurity scattering [16].

Experimental determination of MOSFET channel mobility for incorporation into a model
of drain current usually follows the “effective mobility” technique. Alternatively, estimates can be obtained from measured drain current. In the sections that follow, both approaches are discussed and improvements proposed. For each approach the trap-free case is discussed first so that complications arising from the trap-affected case are more clearly distinguished. In the case of very high interface state density standard techniques are ruined, while when it is low a value close to the real result can be obtained. As discussed in the trap-affected sections the demarcation between what is “high” and what is “low” is roughly given by the magnitude of \( qD_u \) relative to \( C_{ox}' \). Chapter 5 discusses a technique that can be resorted to for the high \( D_{it} \) case.

### 4.1 Discussion of the mobility measurement from C-V

Steady-state MOSFET drain current is given in equation (3.2) in terms of the position-dependent portion of the minority carrier quasi-Fermi level. The classical position-dependent inversion charge density is either computed by the charge sheet model or more exactly as a numerically evaluated integral of minority carrier density throughout the bulk. The mobility outside the integral is labeled with subscript ‘e’ to indicate that it is an effective value given by \( (\int \mu Q'_i dV)/(\int Q'_i dV) \).

\[
I_{ds} = -\mu_e \frac{W}{L} \int Q'_i dV \tag{4.1}
\]

Effective mobility can be isolated from (4.1) leaving drain current and an integral of charge density throughout the channel as the remaining quantities to be determined. In the limit as \( V_{DS} \to 0 \) the integrand approaches a constant equal to its source-side value. The channel be-
comes uniform and $\mu_e$ approaches $\mu$.

$$
\mu = \frac{L}{-WQ_{i0}} \left( \frac{I_{ds}}{V_{DS}} \right)_{V_{DS} \to 0}
$$

(4.2)

Source-side inversion charge density can be measured indirectly using the split C-V technique, which was originally utilized for measurement of interface states [54]. For split C-V the bulk response is separated from that of the channel. The channel capacitance is measured with the bulk grounded and measuring the capacitance between the source/drain and the gate. This capacitance density is a measure of the change in inversion charge density owing to a change in the gate voltage. The channel capacitance is often referred to as the “gate-to-channel” capacitance, which would imply a capacitance variable with subscript “GC.” However, following convention the subscript order should be reversed [28].

$$
C'_{CG} = -\frac{dQ'_i}{dV_G}
$$

(4.3)

Therefore, a measurement of $C'_{CG}$ can be integrated to obtain inversion charge density. Since $V_{DS} = 0$ in the split C-V method the inversion charge density resulting after integration is equal to $Q_{i0}$ contained in equation (4.2). Approximation is indicated in the following equation since discrete integration contains some degree of inaccuracy.

$$
Q_{i0}(V_g) \approx -\int_{-\infty}^{V_g} C'_{CG} dV_G
$$

(4.4)

The other remaining term in (4.2) is the ratio of drain current to drain voltage, $(I_{ds}/V_{DS})_{V_{DS} \to 0}$.
which is usually referred to as channel conductance. Since channel conductance cannot be measured exactly at the DC bias \( V_{DS} = 0 \), the proper approach is to measure the ratio at each gate bias for multiple values of small \( V_{DS} \). A linear fit to the set of \( I_{ds}/V_{DS} \) values gives an equation with y-intercept equal to the ratio’s value as \( V_{DS} \to 0 \) [55, 56]. This is the correct approach implied by equation (4.1).

An incorrect approach is often followed by investigators not familiar with the details of mobility extraction. It is not unusual to find reports where the channel conductance has been improperly determined, which is acceptable for transistor-to-transistor comparison but not for drain current modeling. One mistake is to simply take a single ratio of \( I_{ds}/V_{DS} \) at a very small \( V_{DS} \) value. The smallest \( V_{DS} \) imaginable would technically be acceptable if it were not for noise limitations in real life associated with the measurement equipment. Another mistake is to interpret \( I_{ds}/V_{DS} \) as the slope of drain current measured as a function of drain voltage for some small positive range of drain voltages. This produces a “channel conductance” in disagreement with the outcome of the correct procedure involving extrapolation of \( I_{ds}/V_{DS} \). These arguments can be proven by simulations with either the Pao-Sah expression for drain current or equation (4.1) evaluated with inversion charge density obtained by the charge sheet approximation.

Figure 4.1 shows how a finite drain-source voltage in the current measurement produces an error in extracted mobility following the incorrect approach. Results were calculated for a fictional device by numerically integrating (4.1). \( C'_{CG} \) is given by \( C'_{i}/(1 + C'_{s}/C'_{ox}) \). The mobility-field relationship was assumed to be \( \mu_0/(1 + E_e^2/K^2) \) with \( K = 1 \) MV/cm, \( \mu_0 = 100 \) cm\(^2\)/Vs. The obvious conclusion from Figure 4.1 is that incorrectly utilizing the effective mobility technique yields incorrect conclusions and lays the foundation for incorrect mobility models. It has been shown that data obtained while correctly measuring mobility [57] is quite different (at low vertical fields) from the outcome of the incorrect approach [58, 59].
In the following subsection it is shown that it is possible to derive a correction for the incorrect approaches using long-channel MOSFET theory. It is particularly geared for the incorrect approach that involves measurement of a single ratio of $I_{ds}/V_{DS}$ at a very small $V_{DS}$ value, but is also capable of correcting the slope-based approach by choosing the maximum $V_{DS}$ used. The resulting expression can be used to investigate older mobility results that may have been analyzed incorrectly. In order to compare the quality of the method, results were digitized from a published work containing a modification of the split C-V measurement.

Before beginning an important distinction must be made. Proper analysis using extrapolation implied by equation (4.2) is intrinsically an attempt to estimate channel mobility $\mu$. In contrast, effective channel mobility, $\mu_e$, can be found by measurement at finite $V_{DS}$ (assuming some form of correction discussed in the next subsection). The difference between the two quantities is present because $\mu_e = (\int \mu Q'dV)/(\int Q'dV)$. This relationship reduces to $\mu$ in the limit as $V_{DS} \to 0$. The difference between $\mu_e$ and $\mu$ tends to be less than 1% at small $V_{DS} \leq 100\text{mV}$ (when $\mu_e$ is found correctly), but making the distinction is still important for theoretical clarity. For this reason the subsection to follow is technically a discussion of $\mu_e$. 
Figure 4.1: Actual mobility and equation (4.5) applied at three different drain biases. All curves tend towards the same limiting value at high gate bias but include some artificial roll-off near threshold (around 0.6V). $W = L$, $N_A = 10^{16}$ cm$^{-3}$, $t_{ox} = 200\AA$, $T = 25^\circ C$, trap-free (same parameters used for any other fictional device results to follow).
4.1.1 Analytical correction for the trap-free case

In the early 1990’s C.L. Huang investigated the effects of finite drain bias on effective mobility extraction. At that time it was noted that single point measurements of drain current were commonly performed with $V_{DS}$ values between 40 and 100 mV [60]. It was suggested that to avoid inaccuracy approaching threshold one should use either a somewhat-complex correction [61] or change the experimental capacitance setup [60]. The latter applied the finite drain bias used in the drain current measurement step to the capacitance measurement step. This is necessary to maintain the same channel conditions between the current and capacitance measurements when not performing the necessary extrapolation implied by equation (4.2). Failing to match channel conditions and using the result anyway gives an incorrect effective mobility here labeled $\mu^*_{e}$.

$$\mu^*_{e} \approx \frac{I_{ds}L}{WV_{DS}Q_{i0}}. \quad (4.5)$$

Equation (4.5) is linked to equation (4.1) by the assumption that $\int Q'_{i}dV \approx V_{DS}Q_{i0}$ where $Q_{i0}$ is found from the integral of channel capacitance density, $C'_{CG}$. This capacitance density is typically measured with $V_{DS} = 0$, which corresponds to a different channel condition than that created during current measurement where $V_{DS} \neq 0$. Consequently, equation (4.5) contains an over-predicted inversion charge density and artificially lowers calculated mobility.

For the purposes of developing an analytical correction to equation (4.5) the necessary goal is to improve the approximation $\int Q'_{i}dV \approx V_{DS}Q_{i0}$. Since the Pao-Sah description of drain current has been found to be an excellent descriptor of long channel device behavior, bias point simulations of the $Q'_{i}, V$ relationship point to the needed improvement. For small drain bias $Q'_{i}$ in the channel has a linear relationship with respect to imref splitting, $V$. The mean value
theorem therefore indicates that \( \int Q'_i dV = V_{DS}(Q_{i0} + Q_{iL})/2 \).

Since only \( Q_{i0} \) is measured in the standard split C-V technique some correction is in order assuming the linear \( Q'_i \) assumption is correct. Re-writing the integral \( \int Q'_i dV \) as \( (Q_{i0} + \Delta)V_{DS} \) then \( \Delta = (Q_{iL} - Q_{i0})/2 \). The difference between drain- and source-end inversion charge densities can be found by evaluating equation (3.6) at both locations and taking the difference.

\[
\Delta = C'_{ox}/2 \left[ \psi_{sL} - \psi_{s0} + \frac{\sqrt{2q\varepsilon_sN_A}}{C'_{ox}} (\sqrt{\psi_{sL}} - \sqrt{\psi_{s0}}) \right] \quad (4.6)
\]

Equation (4.6) can be approximated by using the observation that \( \psi_{sL} - \psi_{s0} \approx V_{DS} \frac{C'_{CG}}{C'_{ox}} \). The observation that certain quantities track \( C'_{CG} \) was first reported by C.G. Sodini (albeit for lateral field and inversion charge density) based on two-dimensional MOSFET simulations as part of generating an equation for the extraction of \( \mu_e \) independent of threshold [62]. The difference of square roots is approximated by taking a first order expansion of \( \sqrt{\psi_{sL}} \) about the point \( \psi_{s0} \). In this case the difference of square roots simplifies. For the silicon substrate case in strong inversion one can often assume \( \sqrt{\psi_{s0}} \approx 1 \). For other materials the value of \( \psi_{s0} \) can be replaced with its “pinned” strong inversion value \( 2\phi_F + n\phi_t \), where \( n \) is some integer often suggested to be six. Alternatively, in this work \( \psi_{s0} \) was estimated as \( E_c/q - E_F/q \), which can be approximated as \( E_g/2q + \phi_t \ln(N_A/n_i) \). When these approximations are in error the mobility extraction generally is not possible anyway (subthreshold region discussed next).

\[
\Delta \approx V_{DS}C'_{CG}/2 \left[ 1 + \frac{\sqrt{2q\varepsilon_sN_A}}{2C'_{ox}\sqrt{\psi_{s0}}} \right] \quad (4.7)
\]

The quality of \( \Delta \) deteriorates near threshold for higher \( V_{DS} \), shown in Figure 4.2b. Eventu-
ally the close fit demonstrated in Figure 4.2a cannot be realized if $V_{DS}$ is raised high enough since $\psi_{sL} - \psi_{s0}$ will no longer track $C_{CG} V_{DS}$. Near- and below-threshold correction is difficult for three reasons: 1) unless $V_{DS}$ is very small, near 10 mV, then channel inversion charge density is no longer linear and the mean value theorem no longer gives an obvious result; 2) the $\Delta$ correction eventually fails in subthreshold anyway because $\psi_{sL} - \psi_{s0} \approx V_{DS} \frac{C_{CG}}{C_{ox}}$ degrades; 3) subtraction of overlap and fringing fields that are inevitably present in a real measurement limit the starting point of $C_{CG}$ integration. Because of these reasons it is easy to overpredict or underpredict inversion charge density and insert non-physical behavior in extracted mobility. However, the $\Delta$ correction permits mobility analysis closer to threshold than equation (4.5) and therefore lower in vertical field.
Figure 4.2: Quality of $\Delta$ for two small $V_{DS}$ values as function of $V_{GS}$. (c) indicates threshold.
Finally, an improved estimate for the integral of channel inversion charge density with respect to $V$ can be obtained. The integral of $C'_{CG}$ estimates $-Q_{i0}$ and is indicated as an approximation since discrete integration of a non-linear function is imperfect.

$$\int_{V_{SB}}^{V_{DB}} Q'_i dV \approx (Q_{i0} + \Delta)V_{DS} \quad (4.8)$$

$$\approx -\left(\int_{-\infty}^{V_g} C'_{CG} dV_G - \Delta\right)V_{DS} \quad (4.9)$$

Therefore, an improved estimate for effective mobility using a single $V_{DS}$ point can be realized without experimental modification assuming doping and oxide capacitance are known. $I_{ds}, C'_{CG},$ and $\Delta$ all are gate bias dependent. If discrete integration of $C'_{CG}$ begins at the $i$-th index of $V_G$ then the following equation is valid from $i + 1$ onward (the integral of $C_{CG}$ at the initial point is set to zero).

$$\mu_e \approx \frac{I_{ds}L}{WV_{DS}(\int_{-\infty}^{V_g} C'_{CG} dV_G - \Delta)} \quad (4.10)$$

Although a useful observation from Sodini [62] was used to develop the analytical correction his expression for mobility was not used because it overestimates the integral of channel inversion density same as (4.5). Sodini did account for a diffusion component of drain current and his expression improves the extraction slightly, but the drift-diffusion form of drain current is not necessary to work with in the first place. The drift-diffusion form of $I_{ds}$ and the form involving the gradient of imref splitting are written differently but equivalent, at least until the charge sheet approximation is made [43].
\[ I_{ds} = -W \mu \left[ Q'_i \frac{d\psi_s}{dx} + \phi_i \frac{dQ'_i}{dx} \right] \]  
\[ = -W \mu Q'_i \frac{dV}{dx} \]  

In general, equation (4.2) should be used for analysis by measuring current as a function of small \( V_{DS} \) at every gate bias. This has been mentioned before by J. Hauser [63] but is not always followed in published works containing effective mobility results. Recently, Thomas et al. [55, 56] discussed improvements to the Sodini method by using extrapolation of \( I_{ds}/V_{DS} \). The method works well above threshold but assumes the integral of \( C_{CG} \) can be determined precisely in subthreshold, which may be unlikely because of uncertainty introduced after removal of fringing/overlap capacitance. Also, since \( I_{ds} \) vs \( V_{DS} \) is not entirely linear nor exponential in a small region near threshold, any extrapolation could include some error. The expression given by Thomas was based on Sodini’s expression, which as mentioned above was written starting from the drift-diffusion form of drain current. In this work it is proposed to simply work with the gradient of \( \text{imref} \) splitting, in which case the method of Thomas above threshold assumes the form given by equation (4.2).

Around a decade ago S. Takagi proposed a way of correcting for the drain bias discrepancy by changing how \( C'_{CG} \) is integrated [64]. In his work it was also observed that the channel inversion charge density was approximately linear with respect to potential and a modification was proposed that gets closer to the true mobility value. The resulting expression still contains some overprediction depending on the drain bias used:
\[-\int_{V_{SB}}^{V_{DB}} Q_i' dV \approx \frac{V_{DS}}{2} \left( \int_{-\infty}^{V_g} C'_{CG} dV_G + \int_{-\infty}^{V_g-V_{DS}} C'_{CG} dV_G \right). \quad (4.13)\]

Equation (4.13) must be carefully evaluated when working with discrete data. Only the current subinterval at point $V_g$ is modified by the drain bias; the area of all preceding subintervals can be evaluated normally. In order to evaluate usefulness of equation (4.13) the value of $C'_{CG}$ at a particular gate bias minus the drain bias was found by linear interpolation.

Figures 4.3 through 4.5 demonstrate for a fictional device the ability of $\Delta$ to improve mobility extraction when extrapolation or experimental modification is not used. Multiple curves are shown: 1) the exact $\mu_e$, 2) uncorrected effective mobility from (4.5), 3) with analytical correction, 4) Takagi method [64], 5) Sodini result [62]. Current was numerically computed from (4.1). The result given by extrapolation of $\left(\frac{I_{ds}}{V_{DS}}\right)$ with equation (4.2) is not shown since it is practically the same as the analytical correction.

Proven theoretically by Figures 4.3 through 4.5, an analytical correction for single $V_{DS}$ measurements has been introduced that is very accurate outside threshold and gives almost the same result regardless of the drain bias (provided it is roughly no larger than 100mV). It can be used to re-analyze old mobility data that may have been incorrectly calculated. If mobility were initially calculated incorrectly by equation (4.5) giving $\mu_e^*$, then a corrected version is easily obtained.

\[
\mu_e \approx \left( \frac{1}{\mu_e^*} - \Delta \frac{V_{DS} W}{I_{ds} L} \right)^{-1} \quad (4.14)
\]

While the measurement technique of Huang and proper realization of equation (4.2) are
the best approaches to mobility extraction, the derived correction is a useful alternative. In the absence of any correction, using the uncorrected expression \( \mu_e^* \) in (4.5) is improved when lower \( V_{DS} \) values are used, as this allows for viewing mobility at lower and lower vertical fields. Extrapolated threshold in each of the following figures is around 0.6V, which is important since low field roll-off can be observed at \( V_{GS} = 1V \) even when \( V_{DS} = 10mV \). As already discussed, this low field roll-off is artificially created by the extraction methods. Therefore, artificial roll-off must be separated before conclusions are drawn regarding coulomb scattering.

![Graph showing mobility at different \( V_{DS} \) values](image)

**Figure 4.3:** Effective mobility at \( V_{DS} = 100 \text{ mV} \).
Figure 4.4: Effective mobility at $V_{DS} = 50$ mV.
Figure 4.5: Effective mobility at $V_{DS} = 10 \text{ mV}$. 
The analytical correction described by $\Delta$ has been applied on real data contained in [60]. The results reported by Huang were digitized as proof that the analytical correction is valid. The data is ideal for comparison since the work by Huang in [60] measured channel capacitance at the same drain-source bias as used during current measurement. Such an approach gives an exact calculation of effective mobility, $\mu_e$. Since $\mu_e^*$ was also reported, the analytical correction was applied using equation (4.14) and compared to the digitized $\mu_e$.

Some small, unavoidable error is included due to inexact digitization particularly at lower gate voltages, and for this reason no numerical quantification of error is shown. $C_{CG}$ was not shown past $V_{GS} = 2V$ while the drain current was shown up to 5V, so $C_{CG}$ was assumed constant beyond its maximum. As can be seen in Figure 4.6 the analytical correction produces a corrected curve lying very close to Huang’s experimentally corrected curve (all data digitized from [60]). The datapoints corresponding to the solid curve in Figure 4.6a were not reported all the way down to $V_{GS} = 1V$. 

92
Figure 4.6: All data obtained from C.L. Huang (reference in text). The analytical correction works well since doping and oxide thickness were well-known and $D_{it}$ was very low.
4.1.2 Accounting for interface states

The previous subsection assumed $D_{it}$ was negligible or that a high frequency C-V curve could be generated for which interface states were unable to respond. Capacitance contributed by interface states can be considered a parasitic but is treated differently than the contribution from overlap/fringing fields. Assuming a symmetrical device then overlap adds in parallel to the intrinsic capacitance and can be subtracted out. However, the interface state capacitance is not so easily removed in a measurement of $C_{CG}$, which takes on a new meaning since both the inversion layer and interface trapped charge respond to changes in $V_G$.

$$C_{CG} = -\frac{d(Q'_i + Q'_{it})}{dV_G}$$

(4.15)

If measured in addition to capacitance the differential conductance will register as non-zero as a consequence of carrier capture/emission into/from interface states. This forms the basis of the conductance method, which analyzes bias-dependent peaks in frequency [6]. Since the theory behind the conductance method yields bias region specific closed form expressions it is more advantageous to use gated diodes or MOSFETs with the source/drain/bulk all grounded, termed the full conductance method [65]. As pointed out in [65] high trap densities such that $qD_{it} > C_{ox}'$ create significant frequency dispersion to the extent that at first glance the C-V curve falsely implies a frequency dependent flatband voltage. For the case of $qD_{it} > C_{ox}'$ it is not always possible to obtain a high frequency curve where interface states no longer respond and $C_{CG}$ can reduce to $-dQ_i/dV_G$.

Even if interface state densities are low enough to not cause significant frequency dispersion then the integration of $C_{CG}$ to obtain $Q_i'$ is not necessarily straightforward. It so happens that the x-axis (gate voltage) in a capacitance measurement is not immediately suitable for use as
the variable of integration when interface states are present. At each bias point the response of $Q_{it}$ to a rapidly varying gate voltage is frequency dependent and at high frequency approaches zero. Then, $C_{CG}$ becomes the high frequency $C_{CG}^{HF}$ and reduces to $dQ_i/dV_G$. However, the DC x-axis is still stretched-out by interface states (through the SPE) and an incremental change in DC gate bias does not elicit the same response as the time-varying signal applied to the gate by the capacitance measurement apparatus. Let $C_{it}^0$ be the low frequency ($f \to 0$) value, $v_G$ the time-varying gate voltage, and $V_G$ the DC gate voltage. If the period of the time-varying $v_G$ is made sufficiently fast such that interface states cannot respond then the following is true.

\begin{align*}
C_{CG}^{HF} &= -\frac{dQ_i}{dV_G} \\
&= \frac{C_iC_{ox}}{C_{ox} + C_s} 
\end{align*}

(4.16) \hspace{1cm} (4.17)

However, a change in DC gate voltage gives an expression with a slightly different denominator.

\begin{align*}
-\frac{dQ_i}{dV_G} &= \frac{C_iC_{ox}}{C_{ox} + C_s + C_{it}^0} 
\end{align*}

(4.18)

The difference between the above two equations is that a rapidly varying gate voltage removes the interface state response but the change in DC bias point includes a change in interface state charge density (assuming a slow ramp and long hold time). Therefore, $C_{CG}^{HF}$ integrated with respect to a DC change in gate voltage is incorrect without some modification. Unless the bias-dependent semiconductor and interface state capacitances can be modeled then the inversion charge density will be overestimated to a degree dependent upon the magnitude of $D_{it}$. 

95
For very low $D_{it}$ magnitude then no correction is required.

$$- \int_0^{Q_i} dQ_i' = \int_{-\infty}^{V_S} \frac{C_{HF}}{C_{CG}} \frac{C_{ox} + C_s}{C_{ox} + C_s + C_{it}^0} dV_G$$  \hspace{1cm} (4.19)$$

If $qD_{it} < C_{ox}'$ then a Terman-like analysis [66] or applying the conductance method is likely to result in a good estimate for the low frequency interface state capacitance, at least for the capacitance corresponding to the midgap states. The latter method is not ideal for examining interface states near the band edge unless the temperature is scanned since the peaks in $G_p/\omega$ move too high in frequency [65]. If Hall structures are available then correction for $D_{it}$ can be performed [67]. In conclusion, the amount of $D_{it}$ related de-embedding which must take place to use the split C-V method provides the motivation to look elsewhere for the determination of mobility.

### 4.2 Mobility from drain current

Because of the frequency-dependence in C-V due to $D_{it}$ it is sometimes easier to instead analyze DC current characteristics. The major assumption involved in using trap-affected drain current is that $Q_{it}'$ saturates close to the extrapolated threshold voltage. The latter is very important since the well-known triode equation is written without concern for the presence of $Q_{it}'$ in equation (3.6), which if saturated effectively becomes part of $V_{th}$ in the triode equation.

The well-known triode equation for drain current ($V_{DS} \leq V_{GS} - V_{th}$) can be written to include series resistance by noting that the source is $I_{ds}R_s$ volts above its applied DC potential and the drain is $I_{ds}R_d$ volts below its applied value. Assuming a symmetrical device then $R_s = R_d = R_{sd}/2$, and the current in the triode region can be simplified to account for the reduction in
\[ I_{ds} = \mu C'_{ox} \frac{W}{L} (V_{GS} - V_{th} - \frac{\alpha}{2} V_{DS}) (V_{DS} - I_{ds} R_{sd}) \] (4.20)

\( \alpha \) is the coefficient that results after linearization of the bulk charge term, \( Q'_b \). Its value is not important for mobility analysis and can be assumed as unity. The geometries \( W \) and \( L \) are effective values and therefore it is preferable to use devices with long and wide gate peripheries such that the effective values are approximately equal to the drawn values.

Solving for \( I_{ds} \) and replacing \( \mu \) with its field-dependent quantity the reduction in drain current becomes obvious. Altogether, the total reduction in drain current is caused simultaneously by series resistance and effective mobility. Major deviations from linearity are contained entirely in the denominator.

\[ I_{ds} = \frac{\mu_0 C'_{ox} \frac{W}{L} (V_{GS} - V_{th} - \frac{\alpha}{2} V_{DS}) V_{DS}}{1 + f(E) + \mu_0 C'_{ox} \frac{W}{L} (V_{GS} - V_{th} - \frac{\alpha}{2} V_{DS}) R_{sd}} \] (4.21)

Figure 4.7 demonstrates how the ideal current is reduced by the effective mobility expression and series resistances in a similar way (same parameters as in Figure 4.1). Series resistance was accounted for by using a numerical search for the intrinsic drain-source voltage. The modeling strategy will be to fold the effect of series resistances into the mobility degradation expression \( f(E) \).

The form of \( f(E) \) is either empirical or empirical-universal. The latter seeks to model \( \mu \) such that changes in body bias predict the value and decay of mobility in the high field region. Phonon scattering and surface roughness are thought to be relatable to effective field, \( E_e = \)
Figure 4.7: Drain current reduced by series resistance and field-dependent mobility. $R_{sd} = 5k\Omega$. 

$(\eta|Q'_f| + |Q'_b|)/\varepsilon$, and in the case of scattering by surface roughness then $f(E) = (E_e/E_0)^v$. $E_0$ is the high-field roll-off location and is one of the empirical components along with $\eta$ and $v$ which are process and wafer orientation dependent. In the empirical approach $f(E)$ is replaced with a power series in terms of $(V_{GS} - V_{th})$ and in a version of the Berkeley Short-Channel IGFET Model it was set to $\theta_1(V_{GS} - V_{th}) + \theta_2(V_{GS} - V_{th})^2$ [70].

The drawback to using drain current for mobility investigation is that it requires some assumptions regarding $f(E)$ in order to draw physical conclusions. Also, it is fundamentally limited to the super-threshold region since this is where analytical expressions are valid, which means that lower field components of $f(E)$ cannot be observed. In silicon MOSFETs, the total mobility is some combination of mobilities affected by coulomb scattering, phonon scattering, and surface roughness scattering. Based on experiments these have been found for an n-channel MOSFET to yield an $f(E)$ with several coefficients [57].
\[
 f(E) \approx aE^{1/3} + bE^2 + c \frac{Q_b^2}{Q_i^2 + Q_b^2}.
\]

(4.22)

To compute all elements of \( f(E) \) some fitting approach must be adopted that makes use of trap-free split C-V data. A drain current only approach that assumes \( f(E) \) is simply \( 1 + KE^v \) is possible if the substrate bias can be varied [69]. However for single device modeling where reproducing universality is not a concern a relaxed approach can be employed. Adopting a second order power series representation of \( f(E) \) it is possible to analytically obtain expressions that isolate \( \theta_1 \) and \( \theta_2 \).

\[
 f(E) \approx \theta_1(V_{GS} - V_{th}) + \theta_2(V_{GS} - V_{th})^2
\]

(4.23)

The contribution from series resistance in equation (4.21) can be folded into \( \theta_1 \) for analysis. Consequently, the bottom curve in Figure 4.7 can be assumed equivalent to one free of series resistance by modifying the mobility degradation function appropriately. To fully isolate the reduction in current due to mobility from the reduction due to series resistance it is best to fabricate many identical devices with different gate lengths and perform a regression on an appropriate quantity [70].

The numerator of equation (4.21) can be estimated from a linear fit of \( I_{ds}/\sqrt{g_m} \) in its linear region, which also gives \( \mu_0 \) [71]. For a second order power series in \( f(E) \) the quantity \( I_{ds}/\sqrt{g_m} \) is only linear when the square root in the denominator is approximately unity. This region occurs close to threshold outside of high vertical fields (high gate bias). If no high field deviation from linearity is observed in \( I_{ds}/\sqrt{g_m} \) then the \( \theta_2 \) term can be set to zero.
\[
\frac{I_{ds}}{\sqrt{g_m}} = \sqrt{\frac{\mu_0 C'_{ox} W}{L} V_{DS}} \frac{V_{GS} - V_{th}}{\sqrt{1 - \theta_2 (V_{GS} - V_{th})^2}}
\]

(4.24)

If the linear fit of \(I_{ds}/\sqrt{g_m}\) is labeled as \(\bar{I}_{ds}\) then the derivative of \(\bar{I}_{ds}/I_{ds}\) should be linear with respect to \(V_{GS}\) and have slope equal to \(2\theta_2\). With \(\theta_2\) known it can be subtracted from \(\bar{I}_{ds}/I_{ds}\) to realize another line with slope given by \(\theta_1\). Since \(\theta_1\) contains mobility reduction and series resistance the drain current can be modeled as having no series resistance and deviation from linearity only caused by changes in mobility. The following equations summarize the extraction procedure.

\[V_{th}, \mu_0 \approx \text{linear fit to:} \ \frac{I_{ds}}{\sqrt{g_m}}\]

(4.25)

\[\bar{I}_{ds} = \mu_0 C'_{ox} W \frac{V_{GS} - V_{th}}{L}\]

(4.26)

\[\theta_2 \approx \text{linear fit to:} \ \frac{d}{dV_{GS}} \left( \frac{\bar{I}_{ds}}{I_{ds}} \right)\]

(4.27)

\[\theta_1 \approx \text{linear fit to:} \ \frac{\bar{I}_{ds}}{I_{ds}} - \theta_2 (V_{GS} - V_{th})^2\]

(4.28)

To ensure that the developed drain current expression in (3.19) operates smoothly from subthreshold through strong inversion, some transformation of \(f(E)\) is needed. Knowing that mobility reduction is most important in strong inversion when \(Q'_i \approx C'_{ox} (V_{GS} - V_{th})\) then basing \(f(E)\) on \(Q'_i\) computed from (3.7) satisfies the need for a smoothly varying reference quantity. The chosen transformation of \(f(E)\) is as follows.
\[ \theta_1(V_{GS} - V_{th}) \rightarrow \theta_1 \frac{|Q'_i|}{C'_{ox}} \]  
\[ \theta_2(V_{GS} - V_{th})^2 \rightarrow \theta_2 \left( \frac{Q'_i}{C'_{ox}} \right)^2 \]  

(4.29)  
(4.30)

### 4.2.1 Application to simulated and real trap-free data

The method outlined above was inspired by existing DC analysis techniques but takes a slightly different approach in the analysis of equation (4.21). The method is next verified on simulated trap-free data where equivalent oxide thickness and doping are known and parameters are the same as in Figure 4.7. Then, the same experimental data used for verifying the analytical correction is again analyzed.

In Figure 4.8a the drain current is compared to the square root of transconductance for the purpose of estimating threshold and \( \mu_0 C'_{ox} \frac{W}{L} V_{DS} \). Together, these two values form \( T_{ds} \). Then, the numerator of (4.21) can be divided out by way of \( T_{ds} \) (indicated as drain current “fit” in y-axis of each figure), and the derivative of the inverse can be used to determine \( \theta_2 \) in the high field region (Figure 4.8b). Finally, \( \theta_2(V_{GS} - V_{th})^2 \) is subtracted from \( T_{ds}/I_{ds} \) and the linear portion used to estimate \( \theta_1 \).

Figure 4.8 shows the fit that results after the linear extrapolations in Figure 4.8a-c. The degradation in current caused by series resistances has been folded into the effective mobility expression via \( \theta_1 \). Additional improvement can be made with an optimization step, which of course would not be as useful without the direct extraction procedure. The fit is generated by,

\[ I_{ds} = \mu_0 C'_{ox} \frac{W}{L} \frac{(V_{GS} - V_{th})V_{DS}}{1 + \theta_1(V_{GS} - V_{th}) + \theta_2(V_{GS} - V_{th})^2}. \]  

(4.31)
Figure 4.8: Dotted line indicates the linear fit. (a) estimates $\mu_0$ as 102, compared to 100 actual. (b) estimates $\theta_2$ as $6.99 \cdot (10^{-3}) V^{-2}$, compared to $6.94 \cdot (10^{-3}) V^{-2}$ actual. (c) estimates $\theta_1^e$ as $9.55 \cdot (10^{-2}) V^{-1}$, compared to $9.67 \cdot (10^{-2}) V^{-1}$ actual. (d) is the fit after direct extraction with equation (4.31).
The digitized drain current data used in Figures 4.6a-b in conjunction with C-V data give slightly different results when analyzed separately. The reason is likely due to uncertainty in the effective values of \( W \) and \( L \), different from the mask dimensions after processing and also due to overlap. Therefore, the extracted \( \mu_0 \) obtained from \( I_{ds}/\sqrt{g_m} \) is actually an effective value that contains the deviation of \( W/L \) from its assumed/drawn value. Also, some deviation from assumed oxide thickness is included. Therefore, assuming for any device that \( t_{ox} \) is known, uncertainty in \( W/L \) is not deemed important for replicating only drain current since the extracted value of \( \mu_0 \) corrects for the uncertainty.

Drain current analysis using data in [60] gives different values depending on \( V_{DS} \), caused by what it assumed to be inexact digitization. The digitization process introduces uncertainty in data, which may be the reason there is significant noise in the calculation of \( I_{ds}/\sqrt{g_m} \) in Figure 4.9a. No second order \( f(E) \) effects could be observed in Figure 4.9b, so \( \theta_2 \) can be assumed as zero. At \( V_{DS} = 20\text{mV} \), the following values are extracted: \( \mu_0 = 655 \text{ cm}^2/\text{Vs}, \theta_1 = 0.078 \text{ V}^{-1} \). However at \( V_{DS} = 60\text{mV} \) one finds different values: \( \mu_0 = 609 \text{ cm}^2/\text{Vs}, \theta_1 = 0.0761 \text{ V}^{-1} \). Smoothing drain current with a five-point moving average before computation of transconductance improves agreement and results in \( \mu_0 \) estimates at each drain bias that differ by only 6 \( \text{cm}^2/\text{Vs} \). The source data of Figure 4.9 is from [60].
Figure 4.9: The digitized data from C.L. Huang is indicated with markers (reference given in text). The slope of (a) is related to $\mu_0$ and the slope of (b) is $\theta_1$. In (c) the fit using equation (4.21) is compared against the measured data. Because of the noise in (a) at $V_{DS} = 20\text{mV}$ the low field mobility is overestimated, worsening the result in (c).
4.2.2 Accounting for interface states

An interface state density that rapidly increases near the minority carrier band edge delays onset of the point where \( Q'_i \) versus \( V_{GS} \) approaches \(-C'_{ox}(V_{GS} - V_{th})\), which only occurs when \( Q'_{it} \) saturates to a constant that becomes a part of \( V_{th} \). How much near-threshold voltage stretchout is induced is a function of oxide thickness if \( D_{it} \) were to remain constant. When \( qD_{it} \) is much less than the oxide capacitance density then stretchout is minimized and equation (4.21) is useful. However, when the opposite is true \( Q'_{it} \) takes longer to saturate and linearity in triode region \( I_{ds} \) is delayed by an amount such that significant mobility reduction comes into play before linearity can be observed. This directly impacts the ability to perform current analysis using standard analytical expressions for drain current.

From equation (3.6) there are four components of \( Q'_i \) that vary with gate bias: \( V_{GS}, \psi_s, Q'_b, Q'_{it} \). At large enough gate bias the latter three of these components effectively saturate, but the gate bias where the saturated condition is met shifts with the values of \( D_v, D_c, \) and \( b \) in \( D_{it} \). This is extremely important from a parameter extraction viewpoint since equation (4.21) is not immediately valid above threshold, only becoming valid once \( dQ'_i/dV_{GS} \approx C'_{ox} \). If the interface state density near the minority carrier band edge is large enough then it is possible that once \( dQ'_i/dV_{GS} \approx C'_{ox} \) the mobility has begun to significantly degrade. This ruins the \( I_{ds}/\sqrt{g_{m}} \) method for determining \( \mu_0 \).

To draw conclusions about trap-affected drain current it is possible to investigate the behavior of inversion charge density because drain current is always given by \( I_{ds} = \mu \langle Q_i \rangle V_{DS} W / L \), where \( \langle Q_i \rangle \) satisfies the mean value theorem for integrals for \( Q'_i \) integrated with respect to \( V \). Plots are displayed below which show the behavior of each bias-dependent \( Q'_i \) component as interface state density increases. For demonstration \( V_{SB} = V_{DB} = 0 \), deemed acceptable for drawing conclusions about triode region drain current since the mean value of \( Q'_i \) doesn’t
change much for application of small $V_{DS} \leq 0.1$V. Two plots are shown for every set of $D_{it}$, one with the sub-components of $Q'_i$, and another, (a), showing the derivative of $Q'_i$. Plot (a) for all figures can be thought of as transconductance normalized by oxide capacitance and low field mobility value, $\mu_0$. For linear extrapolation of $I_{ds}/\sqrt{g_m}$ to be successful, the peak value of the dotted line must be close to unity. Clearly, very high $D_{it}$ does not meet this condition. Consequently, analysis of $I_{ds}/\sqrt{g_m}$ would underpredict $\mu_0$.

Figures 4.10 through 4.14 show a sequence of “transconductance” plots for changing $D_{it}$. In Figure 4.10 the trap-free device is displayed as a reference, showing how without traps the transconductance rapidly reaches its maximum value. The result of adding a constant midgap density of interface states is shown in Figure 4.11, which only results in a shift of the threshold voltage and still permits linear extrapolation of drain current above threshold. A larger $D_i$ is simulated in Figure 4.12 to demonstrate what happens when $qD_i$ exceeds $C_{ox}'$, showing that it can slightly delay onset of $Q'_i \propto C_{ox}V_{GB}$. Finally, Figures 4.13 and 4.14 consider the case of an exponentially increasing $D_{it}$ near the minority carrier band edge. This severely distorts transconductance because it takes more band bending (and hence greater $V_{GS}$) for $Q'_i$ to saturate.

It is difficult to define threshold if the derivative of $Q'_i$ does not rapidly approach $C_{ox}'$, since a slow increase to $C_{ox}'$ is reflected by curvature in a plot of triode region drain current. When this difficulty is encountered linear extrapolations of $I_{ds}$ do not give meaningful results and another extraction strategy must be adopted. The following chapter discusses a new form of drain current based $D_{it}$ analysis, where it is shown that $D_{it}$ can be extracted first, followed by mobility. This method is suitable for high $D_{it}$ that exceeds $C_{ox}'/q$. 

106
Figure 4.10: Behavior of transconductance and subparts of $Q'_i$ for a trap-free device. Parameters of the simulated device: $t_{ox} = 200\,\text{Å}$, $N_A = 10^{16}\,\text{cm}^{-3}$, $\mu = 100/(1 + E_e^2/E_0^2)$, $E_0 = 10^6\,\text{MV/cm}$, $V_{FB} = -0.5\,\text{V}$.

Figure 4.11: Behavior of transconductance and subparts of $Q'_i$ for a constant $D_{it}$. 
Figure 4.12: Behavior of transconductance and subparts of $Q'_i$ for a constant $D_{it}$ greater than $C'_{ox}/q$.

Figure 4.13: Behavior of transconductance and subparts of $Q'_i$ for exponentially increasing $D_{it}$. 

108
Figure 4.14: Behavior of transconductance and subparts of $Q'_i$ for exponentially increasing $D_{it}$ for different value of $b$. 

\[
\frac{-dQ_i}{dV_{GS}}/C_{ox} = D_i = 10^{12}, \quad D_c = 3 \times 10^{13}, \quad b = 0.05
\]
Chapter 5

Analysis of the insulator-semiconductor interface through drain current

Significant frequency dispersion in MOS C-V data can be indicative of high interface state density. This frequency dispersion differs from that caused by series resistances in the source drain, which lower the peak capacitance in inversion as frequency is increased. Initially labeled a frequency-dependent flatband shift, the effects of large $D_{it}$ were investigated by K. Martens et al who showed that the frequency dependence is caused by $qD_{it} > C'_{ox}$ [65]. When this relationship is satisfied some extreme behavior is exhibited: the low-frequency gate-bulk and gate-channel capacitances become gate-voltage invariant and equal to the oxide capacitance. As the frequency is raised at a particular DC bias point deviation from $C_{ox}$ is observed.

Interface state density is typically analyzed in the frequency domain through capacitance-voltage and conductance-voltage measurements. Of the available techniques in the frequency domain the conductance method is considered the most accurate, which examines conductance data at each bias point over frequency [6]. Peaks in the parallel conductance divided by angular frequency are related to $D_{it}$ and the locations of the peaks in frequency are related to carrier
time constants. To properly make use of the conductance method the temperature must be varied to get a complete picture of $D_{it}$ over the bandgap, a consequence of the dominant carrier type having extremely small time constants near the band edges. This is particularly a problem for wide bandgap devices where the range of time constants involved is quite large [65].

Other capacitance based techniques such as Terman method [72] and quasistatic variants [73] all have some limitation. Each is limited in accuracy near the band edges, which is again a consequence of the time constants involved at these locations. In the Terman method a “high frequency” capacitance-voltage curve must be measured where interface states cannot respond, but the high frequency condition is not certain to be satisfied since at each bias point the inverse of the trapping time constant must be much smaller than the angular measurement frequency. The problem with higher frequencies is that series resistances become important and they must be deembedded completely before further analysis. In a quasistatic variant it is assumed that capacitance can be measured at such a low frequency that interface states follow completely, but not every device is suitable for low frequencies depending on the device area. This is due to measurement limitations where the detected displacement current is proportional to area, signal amplitude, and frequency. Also, small errors in the oxide capacitance density affect the extraction outcome near the band edges [74].

Charge pumping is another analysis technique but is performed by making use of the bulk contact in a MOSFET [75]. A time-varying voltage is applied to the gate and the resulting substrate current can be related to interface states filling and emptying. The charge pumping current is dependent on frequency, gate area, pulse amplitude, and interface state density. In order to position $D_{it}$ within the bandgap three-level charge pumping must be used [76].

To get a more complete picture of interface states their location in energy must be studied. It has been found experimentally that interface states are often U-like in distribution [4, 14]. A rapid change approaching the minority carrier band edge has been shown to have a signif-
icant effect on the subthreshold drain current behavior of transistors in Chapter 3. Therefore, extracted \( D_{it} \) must be positioned in the bandgap. The conductance method requires some extra effort to perform this positioning (in addition to scanning the temperature), and the charge pumping technique requires three signal levels thus becoming experimentally more complicated. Without resorting to immediate optimization some insight into the surface potential can be obtained from low frequency capacitance data.

The total gate capacitance density, \( C_{GB} \), (also equal to the sum of split-CV capacitances) can be easily manipulated to subtract the oxide capacitance density. At low frequencies what is left is ideally the sum of semiconductor and interface state (at \( f \rightarrow 0 \) Hz) capacitance densities. However, this sum can also be re-expressed in terms of potential by differentiating the SPE in equation (2.14). By integrating the result from flatband or another point [54] where \( \psi_s \) is known then the relationship between surface potential and gate voltage can be obtained without knowing doping [73]. This relationship is used by most \( D_{it} \) analysis methods to locate \( D_{it} \) in the bandgap.

\[
d\psi_s = \left(1 - \frac{C'_{GB}}{C'_{ox}}\right) dV_{GB}
\]  

(5.1)

Later in this chapter techniques only using drain current to extract interface state density are presented. These techniques are not to be confused with the DCIV method [77], which is preoccupied with changes in substrate current. The presented current based techniques only work if the oxide capacitance density is known. Also, the methods assume the bulk is uniformly doped and the dopant concentration has already been determined. These are obvious drawbacks but starting estimates for oxide thickness and doping can be obtained by basic analysis of capacitance-voltage data. The attraction of using drain current is that ideally it satisfies the
condition $f \rightarrow 0$Hz and interface states respond completely without the frequency-dependent concerns intrinsic to a small signal capacitance or conductance measurement.

Before the extraction methods are presented some analysis of bulk doping is outlined for trap-free capacitance-voltage data. Although the extraction methods that follow this section are for the trap-affected case, understanding the effects of nonuniformity is useful for interpreting disagreement between measured and modeled C-V curves.

### 5.1 Doping investigation from capacitance-voltage

Donor or acceptor atoms determine the Fermi level in a material and knowledge of their concentration is vital for the calculation of semiconductor charge density. Capacitance-voltage measurements are advantageous because the relationship between potential and charge density is solved for without much effort. As an example, Poisson’s equation is easily solved for a pn-junction with uniform doping. If one of the junctions is heavily doped then the relationship between applied potential and charge is a simple square root of the product of potential with doping and several constants. If one were to evaluate the capacitance from the resulting charge density then it becomes obvious that squaring the inverse of the capacitance and taking the derivative provides the opportunity to isolate doping. In an MOS device this strategy can also give doping provided the material is uniform and the analysis is performed in a subregion between flatband and threshold (so that majority and minority carriers can be ignored). For nonuniform doping a new strategy must be employed.

Intentional nonuniformity in the bulk dopant level can be desirable for suppression of short channel effects [78]. In [15] a retrograde profile was modeled by including a bias-dependent doping that varied with gate voltage. It was an alternative to an analytical result derived for ion-implanted channels that could be applied towards a threshold-based model [79]. However,
if nothing is known about the doping profile before analysis then a generalized extraction tech-
nique is useful. Only then can an appropriate bias-dependent doping function be formulated. 
The following outlines a means of evaluating doping from capacitance-voltage measurements.

Any extracted “bias-dependent” doping from capacitance-voltage is an effective value. As 
the edge of the depletion region extends away from the surface the total charge density con-
tained within the depletion region is related to an integral of the non-constant doping density. 
For any given surface potential and depletion region width an effective doping density can be 
calculated by equating the nonuniform charge density with the uniform result. This effective 
value changes for every surface potential and depletion region width and is appropriate for 
inclusion within a surface potential based model.

The relationship between the extracted, effective doping and the actual doping is complex. 
If one moves far enough away from flatband then the contribution from majority carriers can 
be ignored. In this case the depletion charge density in terms of the effective density is simply 
given by \( \sqrt{2q\varepsilon_s N_A^e \psi_s} \). Equating this to the exact charge density shows that there are two, 
related unknowns. One is the depletion depth at a particular surface potential and the other is 
the doping density as a function of distance from the surface.

\[
-\sqrt{q\varepsilon_s N_A^e \psi_s} = -q \int_0^{W_{dep}(\psi_s)} N_a(y) dy \tag{5.2}
\]

The doping density and the depletion depth are related to a particular surface potential by 
integrals of doping density over the depletion depth. The actual depletion depth is the physical 
value and is different from the effective value that can be calculated from the effective doping 
density.
\[ \psi_s = \frac{q}{\varepsilon_s} \left( W_{dep}(\psi_s) \int_0^{W_{dep}(\psi_s)} N_A(y) \, dy - \int_0^{W_{dep}(\psi_s)} \int_0^{W_{dep}(\psi_s)} N_A(y) \, dy \right) \]  

(5.3)

The above two equations can be easily derived starting from Poisson’s equation with both potential and electric field equal to zero at \( y = W_{dep} \). However, they usually cannot be analytically solved with doping nonuniform density, \( N_A(y) \). Some trial and error simulation can be performed either using these equations or in a 1-D finite differencing approach where various doping profiles are tuned until an agreeable fit is realized. The finite difference technique would be advantageous near flatband where majority carriers contribute slightly to total charge density.

To demonstrate how the effective doping term changes in response to different doping profiles the one-dimensional finite difference method was applied to Poisson’s equation. The linearized Poisson’s equation using the central limit theorem provides a relationship between node potential and node charge density in the form of a matrix. For \( N \) node voltages there are \( N - 2 \) unknowns since the first node’s potential is given and the last is set to zero. Each node voltage corresponds to a location \( y_i \), and there are \( N \) evenly spaced components corresponding to \( y = 0 \) through \( y = y_{\text{max}} \). The maximum distance from the interface is taken to be some large value such that the electric field approaches zero. An example of the coefficients for the \( N = 6 \) case is shown below. Distance \( d \) is the node separation equal to \( y_{\text{max}} / N \).
The charge density away from inversion reduces to \( qN_A (e^{-\psi_i/\phi} - 1) \) at each node (Maxwell-Boltzmann statistics for holes have been used). Solution of the above matrix is only possible in an iterative fashion because of the non-linear charge density. The exponential arising from holes at a particular node is linearized with respect to that node’s potential but evaluated at the previous iteration’s potential. This changes the coefficients matrix that results from linearization of Poisson’s equation using the central limit theorem. Each matrix term that contains a \(-2\) is replaced with,

\[
-2 \rightarrow -2 \frac{qN_A,i d^2}{\varepsilon_s \phi_t} e^{-\psi_{p,i}/\phi_t}.
\]  

The subscript “p” refers to the previous iteration’s potential, and subscript “i” indicates the node position corresponding to the potential multiplying the coefficient after matrix multiplication. The right-hand side is also modified as a result of charge linearization and each \( \rho_i \) becomes,

\[
\rho_i \rightarrow -qN_{A,i} \left( 1 - e^{-\psi_{p,i}/\phi_t} - \frac{\psi_{p,i}}{\phi_t} e^{-\psi_{p,i}/\phi_t} \right).
\]

In the first iteration the matrices are solved with all \( \psi_{p,i} \) set to zero; this is the initial solution.
that becomes the set of $\psi_{p,i}$ needed for the next iteration. Then, the solution is repeated until the potential stabilizes. The potentials that appear as a result of charge linearization refer to values found from the previous iteration. Increasing the number of nodes, $N$, improves the accuracy of the solution. Ultimately the charge density for a particular surface potential is computed by discrete integral of $-qN_A(1 - e^{-V/\phi_t})$ with respect to $y$. By equating the uniform doping charge density expression with the discrete integral the effective doping from finite difference technique is found to be,

$$N_A^e = \frac{\left(\int_0^{y_{\text{max}}} \rho(y) dy\right)^2}{2q\varepsilon_s [\psi_s + \phi_t (e^{-\psi_s/\phi_t} - 1)]}. \quad (5.7)$$

Figure 5.1 shows three doping profiles and the equivalent uniform doping at each bias point. Profile number one is a retrograde profile. Minority carriers were ignored for demonstration. Profile number three is admittedly a fictional case but shown for demonstration.

### 5.1.1 Extracting effective doping from low frequency C-V measurements

The effective doping ensures the correct relationship between surface potential and gate voltage. For this reason it cannot be extracted directly from depletion capacitance since a bias-dependent doping used in the bulk charge would necessitate changes in the associated bulk capacitance to account for $dN_A^e/d\psi_s$. This would occur even if the bias dependent doping was expressed as a function of gate voltage since that also is a function of surface potential. Instead, one can focus on the relationship between surface potential and gate voltage.

At low frequencies the measured gate-to-bulk capacitance approaches the exact derivative obtained from the theoretical model and equation (5.1) then becomes useful. If surface potential has been evaluated in this manner then the bulk charge contribution to the SPE can be
isolated between flatband and threshold assuming interface states are negligible.

\[
-C_{ox}' \left[ V_{GB} - V_{FB} - \int_{V_{FB}}^{V_{GB}} \left( 1 - \frac{C_{GB}'}{C_{ox}'} \right) dV_{GB} \right] = \tilde{Q}'_b
\]  (5.8)

The line over the bulk charge density is there to indicate it is an interpreted value from measurement. It is uniquely valued at each gate bias. The above solution of $\tilde{Q}'_b$ is the actual value of bulk charge density provided the integration was performed accurately. The flatband voltage can be estimated from the minimum of the first derivative of capacitance with respect to gate voltage. Then, an effective doping can be extracted at each gate bias.
\[ N_A^e = \frac{\tilde{Q}_b^2}{2q\varepsilon_s \left[ \psi_s + \phi_t \left( e^{-\psi_s/\phi_t} - 1 \right) \right]} \]  

(5.9)

## 5.2 Drain current based \( D_{it} \) analysis

In this section the drain current is focused upon since its behavior versus gate voltage is strongly influenced by interface states. An interface state density that rises rapidly towards the minority carrier band has been shown to distort transconductance curves as a consequence of x-axis stretch-out (Figures 3.7, 3.8). The stretch-out occurs because of the expansion in the map of surface potential to gate voltage, which is caused by the presence of \( Q'_{it} \) that aids in balancing the system of MOS charges along with semiconductor charge.

The following two techniques assume that oxide capacitance density and substrate doping are both known. The first technique assumes nothing about the trap density distribution and extracts it directly by comparison with a trap-free curve. The second method assumes interface state density smoothly follows a U-like dependence adequately described by equation (2.32). This second method is quite different from other measurement techniques that assume nothing and deduce the energy dependence of \( D_{it} \) from extracted data. The advantage of the technique is that interface state density near the minority carrier band edge is more rapidly investigated without resorting to measurements over temperature and frequency. The disadvantage of both techniques is that minor disagreements between theory and measurement caused by uncertainty in simulation parameters can be erroneously interpreted as contributions from interface states. Current is examined at low drain-source bias so that the assumed behavior of \( Q'_{t} \) is simpler and a numerical evaluation of \( \int Q'_{t} dV \) is possible with few terms.
Figure 5.2: EOT is the equivalent oxide thickness relative to SiO$_2$.

5.2.1 Method 1

Thin oxides give a large oxide capacitance density and therefore better “immunity” from gate voltage stretch-out for a given interface state density. A useful relationship demarcating the boundary for what constitutes high $D_{it}$ in terms of oxide thickness can be easily derived by equating oxide capacitance density with peak low-frequency interface state capacitance density. The peak value of $C'_{it}$ must be used because $C'_{it}$ is bias-dependent and differs from the simple expression $qD_{it}$ near the band edges. Converting $C'_{it}$ to a pseudo-$D_{it}$ after dividing by $q$ allows for the x-axis to be thought of as apparent interface state density. Following this process shows that oxide thickness is inversely related to apparent $D_{it}$. From Figure 5.2, a 20nm thick oxide has oxide capacitance density exceeded by the low frequency interface state capacitance if apparent $D_{it} \approx 10^{12}$eV$^{-1}$cm$^{-2}$. However, a 2nm thick oxide’s $C'_{ox}$ isn’t exceeded until apparent $D_{it} \approx 10^{13}$eV$^{-1}$cm$^{-2}$. These apparent densities can differ from the actual densities depending on the trap distribution parameters (demonstrated in Figure 2.14).
Figure 5.3: Mobility extracted from $I_{ds}/\sqrt{g_m}$ in its linear region. A large increase in $D_t$ near the minority carrier band edge reduces the apparent $\mu_0$. $t_{ox} = 200\,\AA$, $D_t = 0$, $V_{DS} = 10\,mV$.

It is the peak low-frequency interface state capacitance that should be compared to the oxide capacitance density. Large peaks relative to $C'_{ox}$ ruin estimation of $\mu_0$ from linear extrapolation of $I_{ds}/\sqrt{g_m}$. To get an idea how well the $I_{ds}/\sqrt{g_m}$ technique works different values for $D_c$ are inserted into equation (2.32) and the current analyzed at low $V_{DS}$. The results are in Figure 5.3. For demonstration, mobility is set to $\mu_0/(1 + E_c^2/E_0^2)$ same as in Figure 4.1. The general trend is downward as $D_c$ and $b$ increase. Reducing oxide thickness such that $C'_{ox} > C_{it}$ can shift the curves to the right.

The method in this subsection is valid provided that $C'_{ox} > \max(C'_{it})$ and the assumption of saturated $Q'_{it}$ is true. Then, $\mu_0$ can be estimated from $I_{ds}/\sqrt{g_m}$ and the mobility degradation can be estimated as in Figure 4.8. If the interface state charge density is saturated past peak transconductance then a trap-free model can be used to optimize the mobility values consider-
ing data only past peak transconductance.

With mobility analyzed the following step is to compute a direct extraction of $D_u$ followed by optimization. Direct extraction can be realized by mapping to a trap-free drain current model. For every measured current value a trap-free simulated current value can be found (binomial search, for example). The collection of simulated current values will each have a different gate voltage than that corresponding to the measured current values. Labeling the difference as $\Delta V_G$, this quantity is made up of a constant plus some amount of bias dependence attributed to interface states.

$$\Delta V_G = |V_{G,\text{sim}} - V_{G,\text{meas}}|$$ (5.10)

If $V_{DS}$ is small then trap-free and trap-affected currents occur at roughly the same source-side value of surface potential. The degree to which this is true depends on the magnitude of $V_{DS}$. Figures 5.4a-d were created by allowing a trap-affected drain current to range from 1fA to around 1uA. The gate voltage corresponding to a matching trap-free current was then searched for numerically. Afterwards, the source-side surface potentials for the trap-affected and trap-free cases were compared and plotted as a function of drain current. The difference was also calculated; the plots show that at $V_{DS} = 10\text{mV}$ the difference is minimal. As $V_{DS}$ increases so does the error associated with assuming $\psi_{s0}^{\text{trap-free}} = \psi_{s0}^{\text{trap-affected}}$. 

122
Figure 5.4: Validity of result one obtains by assuming trap-free and trap-affected currents occur at the same source-side surface potential, $\psi_{s0}$. Percent error in (a). Simulated device with constant $\mu = 100 \text{cm}^2/\text{Vs}$, $D_t = 1e12 \text{eV}^{-1}\text{cm}^{-2}$, $D_c = 2e13 \text{eV}^{-1}\text{cm}^{-2}$, $b = 0.1\text{eV}$, $N_A = 10^{16}\text{cm}^{-3}$, $t_{ox} = 200\text{Å}$.
So long as $V_{DS}$ is small then the simulated current used for mapping gives the information needed to compute $\psi_{s0}$. Therefore, taking the derivative of $\Delta V_G$ with respect to source-side surface potential gives an estimate for $C'_{it}$.

$$C'_{it} \approx \frac{d\Delta V_G}{d\psi_{s0}} \tag{5.11}$$

How close one gets to the actual $C_{it}$ is determined by the drain bias. Figure 5.5 continues the process begun in Figure 5.4 and compares the final result to the actual low frequency interface state capacitance. As the drain bias is increased some deviation from the actual value can be observed. Since $C'_{it}/q \neq D_{it}$ near the band edges (as can be seen by the peak of Figure 5.5 not reaching $D_c = 2e13$), the next step would be to fit the extracted $C'_{it}$. The estimate for $D_i$ can be read off from the lowest value of extracted $C'_{it}/q$, although this value may not be possible to read depending on how low in current one can observe and how quickly $D_{it}(E)$ decays to its midgap value.

The above process gives the starting values for $D_i, D_c, b$ and can be thought of as a direct extraction step even though the method is not totally in terms of explicit analytical formulas. Finally, some form of optimization can be performed on each of the three trap values using the chosen method of evaluating drain current. The method is limited by how low in drain current one can reliably measure. Drain-source leakage through junctions can limit the lowest observable drain current. Also, the method assumes that mobility can be adequately modeled before computation of $\Delta V_G$. This assumption is the motivation for exploring another method contained in the next subsection.
Figure 5.5: Method 1 operated on a simulated device. Same parameters as in Figure 5.4.
5.2.2 Method 2: arbitrary $t_{ox}$

When mobility cannot be adequately determined from drain current extrapolation techniques then an alternative is needed since Method 1 discussed in the previous subsection is no longer useful. Taking the ratio of transconductance to drain current at low $V_{DS}$ is such an alternative provided mobility is approximately constant, a condition reasonably satisfied by only analyzing data in and near subthreshold. In subthreshold $g_m/I_{ds}$ becomes “mobility-less” because the rate of change of inversion charge density with respect to gate voltage is much greater than any mobility variation. It is also independent of $W/L$.

To outline clearly the components of $g_m/I_{ds}$ let drain current be written generally as the product of $\mu_e$, $W/L$, $\langle Q_i \rangle$, and $V_{DS}$. The variable $\langle Q_i \rangle$ satisfies the mean value theorem for integrals for the case of $Q'_i$ integrated with respect to channel potential $V$. The ratio of transconductance to drain current is labeled $M$ and is equal to a derivative of the natural log of drain current. It can be evaluated discretely.

\[
M = \frac{d\ln(I_{ds})}{dV_{GS}} \quad (5.12)
\]

\[
\approx \frac{d\ln(\langle Q'_i \rangle)}{dV_{GS}} \quad (5.13)
\]

The approximation in equation (5.13) neglects any contribution from $d\ln(\mu_e)/dV_{GS}$. Assuming the derivative of $\ln(\mu_e)$ is approximately zero is a powerful tool since it is satisfied even if mobility changes slightly. This is because $d\ln(\langle Q_i \rangle)/dV_{GS}$ is very large in subthreshold and only becomes comparable to $d\ln(\mu_e)/dV_{GS}$ above peak transconductance.

Equation (5.12) presents the opportunity to find a set of device parameters using optimization. Components of $D_{xy}$ in equation (2.32) can be determined by optimization if oxide ca-
 capacitance density is known, but starting values are needed. The oxide capacitance density can be estimated from capacitance-voltage, and after optimization the resulting parameters can be used to generate a theoretical C-V for comparison. The optimization can be performed relative to a measured $M$ and a simulated $M$, both calculated discretely using the $V_{GS}$ values from measurement. Starting values can be obtained from Method 1.

The trap-affected simulated current used for simulation of $M$ can be computed numerically (from slowest to fastest) using the Pao-Sah double integral of equation (3.4), the integral form of (3.2) with charge sheet approximated $Q'_i$, or from derived compact drift-diffusion equation (3.20). The Pao-Sah double integral of equation (3.4) is the most physical of the classical cases but the slowest. Computation speeds up drastically for small $V_{DS}$ where linearity in $Q'_i$ versus channel imref splitting is best realized, because fewer points are needed for precise discrete integration. The number of points needed to ensure at most 1% relative error after discrete integration (as compared to many subintervals) is a function of drain bias. For $V_{DS} = 100\text{mV}$, it was found that $N = 4$ subintervals are needed for Simpson’s rule to give sub-1% relative error; for $V_{DS} \leq 60\text{mV}$ only $N = 2$ is needed. These observations permit usage of numerical integration for drain current as a part of an optimization scheme without enduring extremely long run times.

After optimization the final parameter values can be used to form a mobility-free quantity labeled $I_{ds}^0$ that includes voltage stretch-out described by the empirical three-parameter $D_{it}$ model. Then, the effective mobility can be deduced from the measured drain current divided by $I_{ds}^0$. This effective mobility compensates for any error in gate width and length.
\[ \mu_e = \frac{I_{ds,\text{meas}}}{I_{ds}^0} \]  
(5.14)

\[ I_{ds}^0 = -\frac{W}{L} \int Q_i' dV \]  
(5.15)

Method 2 differs from Method 1 in that the former moves quickly into optimization. Method 1 predicts the low frequency interface state capacitance with accuracy dependent upon the value of \( V_{DS} \) used to take drain current measurements and the quality of the mobility extraction. Both require data in subthreshold and near-threshold since this is where the effect of interface states are most noticeable. Since Method 2 is more independent of mobility, Method 1 can be viewed as a means of obtaining initial estimates for portions of the empirical \( D_i \) expression in equation (2.32).

Optimization in this work is performed using the built-in Matlab routine \texttt{lsqnonlin}, which performs nonlinear least-squares analysis. Initial estimates assumed close to the final values are given with appropriate bounds. The convergence criterion is based on when the difference in natural logs of measured and simulated drain current is below some defined tolerance. The initial estimates are crucial to avoid optimization terminating in a minimum corresponding to incorrect model parameters.

One published report was found that is similar to Method 2. The work by H. Katto took ratios of measured transconductance before the peak transconductance point in order to cancel out the approximately constant mobility [80]. The result was compared to a theoretical transconductance and the associated voltage stretchout was related to the surface state density. This approach gives one value for the surface state density and does not attempt to fit the distribution of states near the band edge. Also, Method 2 is very similar to subthreshold slope based techniques but is more general since an average value in the channel is computed.
over all regions of bias. Furthermore, Method 2 does not make the assumption that \(qD_{it} = C'_{it}\) when carrying out the differentiation in equation (5.13), a characteristic common to traditional subthreshold slope analysis [1].

Based on experimental data it was found that simultaneous optimization of flatband voltage, doping, \(D_c\), and \(b\) produced a result indicating approximately constant mobility in subthreshold. The contribution from \(D_i\) was found to be negligible when using drain current, which makes sense assuming \(D_c\) is an order of magnitude or greater than \(D_i\) and that detectable drain current only corresponds to a small range of surface potentials for which the \(D_c\) term dominates. It was found that attempting to set flatband voltage to a fixed value before optimization ultimately gave large mobility variation post-optimization using equation (5.14).

**Quality of measured data and optimization success**

The correct values can be correctly settled upon to a degree dependent on the validity of the \(D_{it}\) distribution assumption (here, U-like) and on measurement noise. Assuming the U-like distribution is correct then small variations in measured current can affect the optimization outcome. The small variations could be attributed to various sources of noise caused by the random motions of carriers in the device itself or the measurement apparatus. Optimization on data containing these small variations can give parameter values that are slightly in error causing the final mobility estimation to also contain some error.

Noise was created using a random sequence of integers that were then scaled to give \(\pm 1.5\%\) variation about unity. Current was multiplied by the resulting dataset (inset of Figure 5.6a). The current was calculated for a simulated device with known oxide thickness, flatband voltage, interface state density, and bulk doping. The last four values were treated as unknowns with a set of starting coefficients different from the actual values. Knowing oxide thickness is justified by assuming one can compare the resulting simulated gate capacitance with a measured gate
capacitance and comparing the peak value in strong inversion. Optimization using $M$ was then performed on the noise-affected current and the resulting parameters compared to the starting values by calculating percent error. The number of current datapoints was kept constant for each on-off ratio.

The percent error in each variable is calculated as a function of the on-to-off current ratio. The “on” current was chosen to be the value at the location of peak transconductance. Figure 5.7 is not universal and is a function of the random noise, meaning its minimum does not represent a universal on/off ratio that is best suited for optimization. Instead, the purpose of Figure 5.7 is to show how measurement uncertainty can alter the extracted result. However, based on these results it is claimed that Method 2 is estimated to give parameter values within 30% of the true values. Improvement can be made by measuring many times to include error bars in the final data and by increasing the number of datapoints.
Figure 5.6: Inset of (a) is the random sequence of integers scaled to give plus or minus 1.5% variation. The noise is not apparent on the plot in (a) until calculation of discrete derivative used for quantity $M$. (b) shows optimization outcome at $I_{on}/I_{off} = 5000$ where the “measured” symbols correspond to simulated current generated for (a).

Figure 5.7: Method 2 and measurement uncertainty.
5.3 Application of methods to silicon MOSFETs

Measurement data corresponding to bulk silicon MOSFETs were obtained from B. Lee [81]. The devices were ideal for analysis since the insulator was grown by atomic layer deposition and verified with cross-sectional TEM giving an excellent estimate for $C'_{ox}$. Also, n- and p-channel MOSFETs with both SiO$_2$ and high-K dielectrics were created giving an opportunity to compare $D_{it}$ as a function of insulator type and doping.

The available device data corresponds to devices with very low EOT on the order of several nanometers, which is beneficial for initial mobility analysis assuming $D_{it} < C'_{ox}/q$. However, thin oxides bring into play quantum confinement of minority carriers, a phenomenon ignored often for thick oxides and low bulk doping where the surface field is smaller in strong inversion. As discussed in relation to Figure 2.4 quantum considerations modify the classical inversion region surface potential and contribute some bias dependent slope in strong inversion. On a plot of capacitance-voltage this means that the peak capacitance in strong inversion is some fraction of the oxide capacitance. Since surface potential is so important to the calculation of interface trapped charge density it is critical to account for quantum mechanical effects when positioning $D_{it}$ within the bandgap. For these reasons the set of equations related to equation (2.16) are used.

An estimate for the EOT of each device was obtained from channel capacitance taken at 100kHz. Since the devices did not suffer from severe gate voltage stretchout a trap free model was used for the capacitance comparison in strong inversion. EOT was modified until the simulated peak capacitance density aligned with the measured peak capacitance density. The initial estimate for doping was obtained from the split C-V bulk response.

Application of Method 1 gives starting estimates for $D_i$, $D_c$ (or $D_v$), and $b$. However, since drain current deep in subthreshold tended to be covered up by drain-source leakage the
Method 1 $D_i$ value is usually suspect. Next, Method 2 was applied using the starting values for flatband voltage, doping, $D_c$ (or $D_v$), and $b$. Depending on the drain-source leakage it is possible for $D_i$ to have no effect on Method 2 and it was excluded for the measured devices. $D_i$ is a midgap value and can have little effect on the SPE when $C'_{ox}$ is large and a slowly decaying $D_c$ component exists.

In total, five silicon MOSFETs were analyzed with Method 2 at $V_{DS} = 100\text{mV}$. The measured and modeled $M$ values are displayed for each device in Figures 5.8 and 5.9. Effective mobility for all devices is plotted in Figure 5.10a versus the absolute value of gate-source voltage and is plotted again in Figure 5.10b versus surface field. The surface field was chosen as the x-axis instead of effective field since technically the form of the latter must be deduced from measurements at different body biases. Mobility at low fields contains deviation attributed to some shortcoming of the assumed $D_{it}$ distribution and measurement noise.

The interface state density distribution implied by outcome of Method 2 is shown for four of the devices in Figures 5.11a-c. The depth into the bandgap that each distribution is shown is dependent on the lowest useful value of current. Subthreshold values of $I_{ds}$ were only considered once they were at least an order of magnitude greater than the junction leakage level. The combined plot in Figure 5.11c demonstrates how little of the bandgap is visible from drain current alone.

The $D_{it}$ that corresponds to the fifth device, a pMOSFET with SiO$_2$ as the insulator, is not shown in Figure 5.11a. Application of Method 2 resulted in $D_v$ and decay parameter $b$ at what were decided to be unrealistic values that were still necessary for a good fit. These values accounted for a slight amount of voltage stretchout in subthreshold on the order of 20mV, which was attributed to compensation for minor inaccuracy in the QM treatment of hole inversion charge density.

Sources of inaccuracy in the Method 2 optimized parameters are from the triangular ap-
proximation in the QM treatment of minority carriers, drain-source leakage currents limiting the view of subthreshold, and measurement noise. These limitations imply that using theoretical equations compared against measured quantities can produce inferior results relative to techniques that evaluate $D_{it}$ directly. However, provided the methods making use of the theoretical equations are applied in a standard fashion to multiple devices then the extracted $D_{it}$ will at least usefully compare the devices in relation to one another.

A comment can also be made regarding the EOT. Thin dielectrics reduce the effects of interface states on the SPE because $Q'_{it}$ is divided by $C'_{ox}$. Analysis of thicker oxides would not be quite as sensitive to minor errors in $t_{ox}$ since the effects of $Q'_{it}$ on gate voltage stretchout would be more pronounced. Also, quantum mechanical considerations for carriers could potentially be ignored since the reduction in gate capacitance is minor and the final mobility value can provide any needed compensation.
Table 5.1: nMOS parameter summary.

<table>
<thead>
<tr>
<th>insulator</th>
<th>EOT (nm)</th>
<th>$N_{sub} \times 10^{17}$ cm$^{-3}$</th>
<th>$V_{FB}$ (V)</th>
<th>$D_{c} \times 10^{12}$ cm$^{-2}$eV$^{-1}$</th>
<th>$b$ (meV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>2.65</td>
<td>1.95</td>
<td>-0.437</td>
<td>6.31</td>
<td>123</td>
</tr>
<tr>
<td>HfO$_2$</td>
<td>2.60</td>
<td>3.21</td>
<td>-0.279</td>
<td>12.6</td>
<td>96.2</td>
</tr>
</tbody>
</table>

Table 5.2: pMOS parameter summary.

<table>
<thead>
<tr>
<th>insulator</th>
<th>EOT (nm)</th>
<th>$N_{sub} \times 10^{17}$ cm$^{-3}$</th>
<th>$V_{FB}$ (V)</th>
<th>$D_{v} \times 10^{12}$ cm$^{-2}$eV$^{-1}$</th>
<th>$b$ (meV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>3.43</td>
<td>1.74</td>
<td>0.671</td>
<td>94.4</td>
<td>1.46</td>
</tr>
<tr>
<td>HfO$_2$</td>
<td>1.65</td>
<td>1.63</td>
<td>0.314</td>
<td>9.67</td>
<td>36.1</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>2.62</td>
<td>1.73</td>
<td>0.865</td>
<td>7.02</td>
<td>28.0</td>
</tr>
</tbody>
</table>

Figure 5.8: Measured and fit ratio of transconductance to current for two n-channel MOS-FETs. $V_{DS} = 100$mV.
Figure 5.9: Measured and fit ratio of transconductance to current for three p-channel MOS-FETs. $V_{DS} = 100\text{mV}$.
Figure 5.10: Combined effective mobility implied by results of Method 2.
Figure 5.11: $D_{it}$ distribution implied by Method 2.
Chapter 6

Characterization of gallium nitride
MISFETs

Gallium nitride (GaN) MISFETs with SiO$_2$ as the gate insulator were fabricated by M. Veety as part of research directed towards achieving enhancement mode, four-terminal transistors with p-type substrate and metal gates [82]. The p-type GaN body delays formation of the n-channel and was expected to move threshold well above zero, thus creating a normally off device. Normally off devices are critical for single power supply applications such as logic gates provided there is a suitable gate insulator limiting off-state leakage.

The GaN substrate itself has a host of material properties that theoretically make it an attractive substrate as an alternative to silicon. Of these the most interesting is its high breakdown field enabling improved high power transistors for use in electrical switching systems. Other favorable properties relative to silicon include heightened bulk mobility and large thermal conductivity. Even though the material properties of GaN predict improved performance through several Figures of Merit [83] it remains the work of process engineers to solve the basic problems of GaN MISFET creation. These basic problems center around the non-native
insulator/semiconductor interface, source/drain formation, and bulk contact.

The devices analyzed in this work primarily suffer from high interface state density and large series resistances. The latter directly degrades current as demonstrated in Figure 3.11. The former produces significant gate voltage stretchout since thick insulators were used and broadens the sub- to super-threshold transition that becomes most apparent in a plot of transconductance. Both effects combine to yield an apparent reduced mobility (when performing triode region linear extrapolations) and complicate modeling significantly. The modeling is also complicated by the lack of a good ohmic contact to the bulk, which makes substrate doping and flatband voltage impossible to analyze directly from C-V. Attempts can still be made to determine their values from other techniques that will be discussed.

6.1 GaN-specific modeling concerns

A GaN substrate encourages revisiting some basic modeling assumptions commonly made in the classical solution of Poisson’s equation. More specifically, assumptions regarding dopant ionization and carrier degeneracy. Both have the potential to manifest noticeably in capacitance voltage plots depending on temperature and oxide thickness. Another potential modeling concern, the polarization charge at the GaN surface, can be combined into the flatband voltage [84].

The available p-type dopants in GaN all reside greater than several $kT$ away from the valence band [85]. This is in contrast to the silicon case where boron sits only 45 meV away [27]. From a modeling standpoint this means that $N_A^-$ at flatband is some fraction of the total dopant concentration. Magnesium has been found to be one of the better GaN p-type dopants and sits between 200 and 250 meV relative to the valence band [86]. Consequently, the dopants at the surface become ionized with band bending. This occurs well before inversion but introduces
a “wobble” or deviation in capacitance-voltage around flatband that is normally only seen in silicon provided temperature is low enough [87]. From a drain current modeling perspective incomplete ionization can be ignored provided the flatband voltage is shifted slightly.

Unintentional n-type doping should theoretically be included in the bulk charge density expression used to derive the total semiconductor charge density. Compensation of the p-type dopant is likely due to unintentional incorporation of oxygen during growth of GaN crystals [88]. However, it is likely that simply an effective $N_A$ value should be chosen for modeling purposes unless one can perform temperature studies [89].

GaN has an effective conduction band density of states around an order of magnitude smaller than in silicon [18]. From a classical perspective this indicates that carrier degeneracy is more of a concern. For the same doping and oxide thickness, two devices subjected to the same gate bias in strong inversion will have different surface potentials form if one has a lower value for $N_c$. The one with the lower $N_c$ value has an increased surface potential in order to balance the system of charges. Comparing silicon to GaN not only is $N_c$ to blame for this increase but also the slightly lower relative permittivity of GaN that creates higher surface fields for the same band bending. For these reasons it is argued that the quantum mechanical treatment of minority carriers would become non-negligible much faster than for silicon (if one were to scale the oxide downward).

The above considerations change the form of semiconductor charge density for the case of a GaN substrate. Since thick oxides greater than 50nm were analyzed quantum confinement was ignored. Instead, one could choose to use Fermi-Dirac statistics to describe electron and hole concentrations. Fermi-Dirac statistics have been shown to agree reasonably with the QM analysis (at least better than the Maxwell-Boltzmann description) [32]. If Poisson’s equation is solved with equations (2.3) and (2.6) then the result for n-channel with compensating donors is,
\[ Q'_s = \mp \sqrt{2q\varepsilon_s} \left[ N_v \frac{V_T}{\Gamma_{5/2}} \Delta F_v + N_c \frac{V_T}{\Gamma_{5/2}} \Delta F_c + \psi_s (N_A - N_D) \right] + V_T N_A \ln \left( \frac{1 + g_A e^{(-\psi_s + E_A - E_F)/kT}}{1 + g_A e^{(E_A - E_F)/kT}} \right)^{1/2}. \] (6.1)

\[ \Delta F_v = F_{3/2} \left( \frac{-q\psi_s + E_v - E_F}{kT} \right) - F_{3/2} \left( \frac{E_v - E_F}{kT} \right) \] (6.2)

\[ \Delta F_c = F_{3/2} \left( \frac{q\psi_s - (qV + E_c - E_F)}{kT} \right) - F_{3/2} \left( \frac{-(qV + E_c - E_F)}{kT} \right) \] (6.3)

Equation (6.1) has been given before by others [90, 26] and is easily derived by the same process as equation (2.12). Here it is written differently by treating donors as completely ionized and by writing the expression in terms of energy levels instead of potentials. \( F_{3/2} \) is the Fermi-Dirac integral of order one-half, and \( \Gamma_{5/2} \) is the gamma function evaluated at 5/2. The Fermi-Dirac integral can be evaluated with an approximating function [91]. The Fermi level in the presence of compensating donors and large ionization energy is found from the equilibrium \( \rho = 0 \) condition with \( n \approx 0 \).

\[ E_F = E_v + kT \ln (N_v) - kT \ln \left[ \frac{N_D + \xi}{2} \left( \sqrt{1 + \frac{4(N_A - N_D)\xi}{(N_D + \xi)^2}} - 1 \right) \right] \] (6.4)

\[ \xi = \frac{N_v}{g_A} e^{- (E_A - E_V)/kT} \] (6.5)

With Fermi-Dirac statistics the slope in strong inversion is slightly reduced, related to the increased \( \psi_s \) in strong inversion as shown in Figure 6.1. Figure 6.1 indicates that for thicker
Figure 6.1: Theoretical surface potential curves. $SPE_{MB}$ is SPE with Maxwell-Boltzmann statistics; $SPE_{FD}$ is SPE with Fermi-Dirac statistics. For Si: $N_c = 3.2 \times 10^{19}$ cm$^{-3}$. For GaN: $N_c = 2.2 \times 10^{18}$ cm$^{-3}$. $V_{T0}$ is the long-channel threshold voltage. Due to the strong deviation in $\psi_s$ at $t_{ox} = 10$nm it is likely that quantum confinement should be considered in a device model.

insulators it is acceptable to use equation (2.12).
6.2 Analysis of a three-terminal GaN MISFET

A GaN MISFET with around 100nm of SiO$_2$ as the gate dielectric was measured and found to have an $I_{\text{on}}$ to $I_{\text{off}}$ ratio of around $10^4$. The on/off ratio was limited by drain-source leakage between the ion-implanted n-type junctions. The capacitance-voltage curves demonstrated significant frequency dispersion due to high levels of interface states. The bulk doping was unknown due to lack of a proper bulk contact on the Magnesium doped substrate. The drain-source leakage was either caused by a non-abrupt junction with significant G-R centers in the substrate, or due to the net activated bulk doping being weakly n-type.

Correctly analyzing the interface state density would involve a proper bulk contact and analysis in the frequency domain. Additionally, the temperature would have to be varied over a wide range due to the wide bandgap of GaN. A fitting strategy involving peaks in frequency found from the conductance method would ideally be adopted and used to describe the extracted interface state density. Since this approach could not be used a less rigorous one involving drain current was taken. The extracted data contains a fair amount of uncertainty but gives a starting point for further device characterization.

Supporting theory needed to qualitatively explain measured C-V data is first discussed. Then, Method 2 from the previous chapter is applied to drain current data as a means of estimating the peak mobility, $\mu_0$. Most of the conclusions drawn are admittedly qualitative, limited by the fact that the transistor structure is intrinsically incomplete owing to the bulk contact’s absence.

6.2.1 Supporting theory

Frequency dispersion is caused by series resistances and/or interface states. Interface states are unique in that levels satisfying $D_{it} > C'_{ox}/q$ give an apparent frequency-dependent flatband
shift [65]. Analytical modeling of this behavior in the frequency domain is region-specific in regards to the gate voltage. The analytical $D_{it}$ extraction expressions for the conductance method are the result of simplifying the non-regional equations outlined by E. Nicollian and J. Brews that can be evaluated numerically at any gate bias [6]. Figure 6.2 shows the small signal equivalent circuit after combining a continuum of trapping levels. Each trapping level has a unique small-signal response that gives rise to frequency dispersion. Each element in Figure 6.2 was drawn with a box to indicate an admittance block that has frequency-dependent real and imaginary components.

\[ G_n(E) = \frac{q}{\phi_t} D_{it}(E)c_n n_s (1 - f_o) \]  
\[ G_p(E) = \frac{q}{\phi_t} D_{it}(E)c_p p_s f_o \]  
\[ C_T(E) = \frac{q}{\phi_t} D_{it}(E)f_o (1 - f_o) \]  
\[ f_o = \left[ 1 + e^{(E - E_F - q\psi_s)/kT} \right]^{-1} \]

$c_n (c_p)$ is the electron (hole) capture probability with units volume per second, and $n_s (n_p)$ is the electron (hole) volume density at the surface. These quantities are used to evaluate the elements in Figure 6.2. Band bending is accounted for through occupation function $f_o$. The following equations can be integrated numerically to make sense of measured capacitance data. They are frequency dependent and have real and imaginary components.
Figure 6.2: Equivalent circuit for the small signal response of interface states. $C_c$ becomes $C_i$ and $C_v$ becomes the depletion/majority carrier response for pMOS; the opposite occurs for nMOS.

\[
G_{gr} = \int_{E_v}^{E_c} \frac{G_n G_p}{j \omega C_T + G_n + G_p} dE \tag{6.10}
\]

\[
C_{Tn} = j \omega \int_{E_v}^{E_c} \frac{C_T G_n}{j \omega C_T + G_n + G_p} dE \tag{6.11}
\]

\[
C_{Tp} = j \omega \int_{E_v}^{E_c} \frac{C_T G_p}{j \omega C_T + G_n + G_p} dE \tag{6.12}
\]

Since the measured devices had a poor bulk contact (highly resistive, VB conceptually terminated in an open circuit) the depletion and accumulation responses (equal to $C_v$ in Figure 6.2) were negligible. The conduction band was contacted to through the source/drain minority carrier supplies. Two-thirds of the trapping elements are negligible because of the negligible bulk responses; the inversion response $C_c$ is in parallel with $C_{Tn}$. The situation differs from split-CV where both bulk and channel components follow the time-varying gate voltage but only the channel response is measured. Here, the equivalent gate admittance density has real
and imaginary components, and the equivalent gate capacitance becomes frequency dependent since $C_{Tn}$ is complex valued (and frequency dependent).

\[ Y'_g = \left( \frac{1}{j\omega C'_{ox}} + \frac{1}{j\omega C'_i + j\omega C_{Tn}} \right)^{-1} \]  

(6.13)

\[ C'_g(f) = \frac{\Im[Y'_g]}{\omega} \]  

(6.14)

The above admittance simplifies to the same equations describing the depletion region commonly used for the conductance method, which is only made possible by the poor bulk contact. The low frequency interface state capacitance $C'_{it}$ is from equation (2.42), and $\tau_n = (c_n n_s)^{-1}$ is a time constant that depends on band bending. The capture probability constant is an unknown quantity often assumed to be bias-independent and is extracted from measurement over temperature.

\[ \Im[j\omega C_{Tn}] / \omega = \frac{C'_{it}}{\omega \tau_n} \tan^{-1}(\omega \tau_n) \]  

(6.15)

\[ \Re[j\omega C_{Tn}] = \frac{C'_{it}}{2 \tau_n} \ln[1 + (\omega \tau_n)^2] \]  

(6.16)

At very low and very high frequencies the gate capacitance reduces to two extremes. These extremes differ from the ideal gate-to-bulk capacitance because of the poor bulk contact and are specific to the device under consideration. The integral in the second equation is just low frequency interface state capacitance $C'_{it}$ from equation (2.42).
\[ C'_g(f \to \infty) = \left( \frac{1}{C'_{\text{ox}}} + \frac{1}{C'_i} \right)^{-1} \]  \hspace{1cm} (6.17)

\[ C'_g(f \to 0) = \left( \frac{1}{C'_{\text{ox}}} + \frac{1}{C'_i + \int_{E_c}^{E_v} C_T dE} \right)^{-1} \]  \hspace{1cm} (6.18)

When \( D_i >> C_{\text{ox}}' / q \) the low-frequency gate capacitance-voltage curve approaches a constant value of \( C'_{\text{ox}} \). At very high frequencies at which no traps can respond then the gate capacitance includes no contribution from \( C'_i \) but there is still DC gate voltage stretchout due to the SPE. A frequency between these two conditions will give a leading edge that moves horizontally with changing frequency.

To prepare for qualitative comparison with the measured device results a fictional device is simulated with \( t_{\text{ox}} = 70 \text{nm} \). Only a midgap density is considered; including a \( D_c \) term would elongate the sub- to super-threshold transition. Simulation results are in Figure 6.3 with three frequencies ranging between 1 kHz and 1MHz. The capture probability constant determines where the leading edge at a particular frequency is located on the \( V_g \) axis. Once \( qD_i \) exceeds the insulator capacitance density significant stretchout occurs. The flatband voltage was fixed at -1V but in reality can be anything depending on metal-semiconductor workfunction, ratio of donor-like to acceptor-like interface states, and fixed/trapped oxide charge. The difference in terms of gate voltage between the leading edges at each frequency is a function of \( D_i \).

The peak of each curve in Figure 6.3 can also be lowered by series resistances originating from the source/drain depending on the frequency. The small signal model in Figure 6.2 would have a resistance placed between the CB terminal and ground. Series resistance can be ignored at lower frequencies but eventually becomes important as the frequency is raised. The apparent peak capacitance becomes degraded from its true value.
Figure 6.3: Voltage stretchout and frequency dependence related to interface states. Top dotted line is related to $C'_g(f \to 0)$; bottom dotted line is $C'_g(f \to \infty)$. Solid lines from left to right: 1kHz, 31.6kHz, 1MHz. The title in each plot gives the interface state density relative to the oxide capacitance density. (b) is an extreme case where greater than 15V of stretchout occurs.
6.2.2 Measurement results

The transistors that yielded were all DC FATFETs exhibiting indeterminate extrapolated threshold voltages. The severe elongation of the sub- to super-threshold transition prevents a clear quantification of threshold and is a hallmark of an increasing $D_{it}$ near the minority carrier band edge. Capacitance-voltage echoed this conclusion via pronounced frequency dispersion. As expected the accumulation region could not be observed regardless of measurement frequency and gate bias. DC gate leakage was negligible as expected for a thick insulator and most devices survived at least until +50V; achieving higher voltages was restricted by the measurement setup.

Figures 6.4a-d show capacitance and conductance data taken over a wide range of voltages and frequencies in an attempt to apply the conductance method on a 64x64 um$^2$ GaN MISFET. Figure 6.4b is the measured parallel conductance from which oxide capacitance is deembedded. The result after deembedding is shown in Figure 6.4c. After deembedding, the peak $G_p/\omega$ shifts downward in frequency and was not visible as shown in the rescaled version of Figure 6.4d. Therefore, the extracted $D_{it}$ from the peaks of Figure 6.4b becomes a lower limit and was estimated to be in the $10^{13}$ cm$^{-2}$eV$^{-1}$ range. Changing the temperature would theoretically shift the peaks to a visible frequency location. Moving the stimulus’ signal lower in frequency is impossible due to physical detection limits of the measurement equipment for such a small area device. The only quantitative conclusion that can be made from Figures 6.4a-d is that $D_{it}$ is at least 2 orders of magnitude larger than $C_{ox}'/q$ causing significant voltage stretchout.
Figure 6.4: Capacitance and conductance results over a range of frequencies. (c) is the corrected curve after deembedding oxide capacitance density, and (d) is the rescaled version showing the peak moving downward in frequency. The dotted line in (a) indicates the estimated oxide capacitance.
Measured drain current of the device from Figures 6.4a-d is shown in Figure 6.5 while cycling through three drain biases at 100, 150, and 200 mV. The shown curves correspond to the reverse sweep direction only since the forward direction showed increased stretch-out due to slow insulator states (hysteresis). Ideally, once a large gate voltage is applied the slow insulator states are filled for the duration of the reverse sweep. Analysis was performed on the current using Method 2 with flatband voltage, doping, $D_c$, and $b$ as the optimization parameters. The initial values of each were estimated.

It has been noted that in SiC the density of interface states implied from measurement can be comparable to the three-dimensional density of states (DOS) in the conduction band depending on the dimension of the region over which $D_{it}$ is distributed [92]. Accordingly, the upper limit of integration in equation (2.30) should be increased until the conduction band DOS increases past the observed $D_{it}$. Since it was known the devices had very high $D_{it}$ this viewpoint was adopted. Five parameters were optimized: $N_A$, $V_{FB}$, $D_c$, $b$, and $\Delta E_c$. The last parameter, $\Delta E_c$, was added to $E_c$ during integration of (2.30) and is some small offset. Moving straight into optimization is admittedly non-ideal but reasonable bounds were chosen: doping was assumed to be somewhere near $10^{16}$ cm$^{-3}$; $D_c$ and $b$ were limited in a range similar to that used for analysis of the SiO$_2$ devices; $\Delta E_c$ was kept below 200 meV, which was used in [92].

The final parameter values are not certain since there is some dependence on drain bias. General claims that can be made with some degree of confidence are that doping is in the upper $10^{16}$ cm$^{-3}$ range, that $\Delta E_c$ is near 100 meV, and that low field effective mobility is near 4 cm$^2$/Vs. The parameter estimates are shown in Table 6.1 for three different bias points, and the disagreement is attributed to: 1) measurement variation caused by charge injected into the insulator (discussed below); 2) inability to see deep into subthreshold caused by junction leakage. The fit curves are shown together in Figure 6.6a along with the implied value of $\mu_0$ in Figure 6.6b. The resulting $D_c$ and $b$ values imply that the maximum $C'_{it}$ is two orders of
Table 6.1: Summary of results obtained from Method 2.

<table>
<thead>
<tr>
<th>$V_{DS}$ (mV)</th>
<th>$N_A$ (cm$^{-3}$)</th>
<th>$V_{FB}$ (V)</th>
<th>$D_c$ (cm$^{-2}$eV$^{-1}$)</th>
<th>$b$ (meV)</th>
<th>$\Delta E_c$ (meV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$7.25 \times 10^{16}$</td>
<td>-1.69</td>
<td>$6.92 \times 10^{13}$</td>
<td>140</td>
<td>99.2</td>
</tr>
<tr>
<td>150</td>
<td>$7.93 \times 10^{16}$</td>
<td>-1.68</td>
<td>$6.32 \times 10^{13}$</td>
<td>146</td>
<td>97.4</td>
</tr>
<tr>
<td>200</td>
<td>$7.45 \times 10^{16}$</td>
<td>-1.32</td>
<td>$5.76 \times 10^{13}$</td>
<td>150</td>
<td>89.8</td>
</tr>
</tbody>
</table>

magnitude greater than $C_{ox}'$, which is in agreement with the lower bound estimate on $D_{it}$ made in Figure 6.4.

The ideal curve that would exist (for a constant mobility) in the absence of interface states is shown in Figure 6.7 side-by-side with the measured result. It is slightly misleading since a truly trap-free device would likely have a higher value of $\mu_0$ due to lack of coulomb scattering from charged interface states. Current was simulated up to the same value as the maximum measured value. The flatband voltage for the ideal trap-free simulation was kept the same as the result from Method 2 since only $D_c$ was estimated and $Q_{it00}/C_{ox}'$ for such a case is negligible.
Figure 6.5: Representative GaN MISFET that could survive out to very high gate voltages. From C-V, SiO$_2$ thickness is around 70 nm. The fluctuation in (c) is the result of discrete differentiation, which has “amplified” some minor deviation in (a). The minor deviation in (a) could be attributed to the measurement equipment, poor probe contact to pads, or some small change in the interface properties due to biasing.
Figure 6.6: Outcome of Method 2 applied to the data in Figure 6.5. In (a) the markers represent measured data, solid lines are the fit. In (b) both the lines and markers represent the resulting mobility. $W = L = 64 \text{um}$ and $t_{ox} \approx 70 \text{nm}$. 

155
Figure 6.7: Reconstructed ideal case for device of Figure 6.5 where $D_{it} = 0$. 
6.3 Measurement of a four-terminal GaN MISFET

A second generation of GaN MISFETs were fabricated by M. Veety using a dedicated mask set with test structures for determining the metallization electrical properties (Figure 6.11). The motivation for the dedicated mask set was that a proper bulk contact allows for split-CV and viewing of the total gate capacitance in accumulation and inversion. The ability to observe majority and minority carrier responses yields a more robust extraction.

Test structures were drawn to individually characterize the bulk and source/drain regions. Transistor sizes were chosen to allow for scaling theory to be applied towards removing parasitic elements. Also, RF transistors with ground-signal-ground pads were drawn for measurement at higher frequencies using a network analyzer. Large geometries were chosen to increase tolerance in regards to unavoidable process drift under humid cleanroom conditions. A probeable fourth-terminal was added to each device for grounding the bulk and allowing for alternative $D_{it}$ techniques such as charge pumping.

The finished 4-terminal GaN MISFETs had some issues centered around junction leakage preventing complete characterization. Either the devices unintentionally became accumulation FETs, or high bulk trap densities existed in addition to non-abrupt ion implanted junctions. In the former case the bulk would be weakly n-type due to poor outcome of Magnesium activation anneal and the source/drain junctions would easily conduct in response to an applied $V_{DS}$. In the latter case high bulk trap densities would act as generation-recombination centers in the junction/bulk depletion region and produce leakage current.

Capacitance-voltage results failed to clearly elucidate the problem. If the bulk were indeed weakly n-type then forward bias would correspond to accumulation and decreasing gate bias enough would place the device into depletion. Assuming large $qD_{it} \gg C'_{ox}$ then the minimum of the set of capacitance-voltage curves taken over frequency would have some significant
slope indicative of a depletion curve. Figure 6.8 exhibits no such behavior but this could be due to inability to sweep frequency high enough. Even if frequency could be significant raised the effect of series resistances would distort the curves as is already happening at the 1MHz curve. Estimations solely based on the real part of the gate impedance estimate $R_s$ in the 5k to 10kΩ range. Some additional confidence is lent towards saying the bulk is not p-type based on TLM results for the bulk ohmic contact that produced rectifying behavior.

In either case the end result is the same: the junctions are extremely leaky and prevent the device from completely pinching off. A quick measurement of the transistor's family of curves in Figure 6.9 shows the junction leakage inducing a tilt. Once gate voltage drops below -5V there is no longer any control over $I_d$ even though the channel current might be zero owing to some parallel path corresponding to junction leakage. This parallel path limits success of drain current fitting, because information over at least several orders of magnitude is more desirable for a successful extraction.
Figure 6.8: Real and imaginary components of the gate small-signal response. 20 frequencies evenly spaced logarithmically from 1kHz (left-most curve) to 1MHz (right-most curve). Series resistances reduce the peak capacitance at high frequencies. Representative plot of typical device.
Figure 6.9: The family of curves are tilted upward by the presence of junction leakage. Representative plot of typical device.
Figure 6.10: Log and linear scale plots of drain current showing how $I_{on}$ to $I_{off}$ is not much greater than a decade. Representative plot of typical device.
Figure 6.11: One quadrant (repeated in three other quadrants) in a 1 cm x 1 cm square. The blue layer is final metal layer used for probing and understandably dominates the layout. Red indicates ion-implant while grey corresponds to mesa isolation. Not visible due to final metal layer are the n-ohmic window, p-ohmic window, and via.
Chapter 7

Conclusions and suggested work

The effect of interface states on transistor performance cannot be ignored for alternative insulators and substrates. In order to extract physical parameter values giving agreement with drain current, it is necessary to include an energy-dependent distribution of interface states. Ignoring the distribution of interface states otherwise results in parameters only valid over localized regions of gate bias. In this work the U-like distribution assumption was selected and modeled using a pair of exponentials.

A characteristic of this work in contrast to others is that no assumption was made regarding the acceptor- and donor-like nature of interface states. As shown in Chapter 2 the bias-dependent nature of interface trapped charge density $Q_{it}'$ is wholly contained by simply using the sum of donor- and acceptor-like trap densities. It has been concluded that distinguishing between the two types is not important for reproduction of drain current as long as the components of flatband voltage are analyzed while acknowledging the presence of the bias-independent portion of interface trapped charge density.

A compact, trap-affected model has been presented. It is physics-based and founded on a surface-potential view of charges in the semiconductor. It functions reasonably well consid-
ering the challenges imposed by the bias-dependent non-linearity of inversion charge density with respect to surface potential. The compact expression is expected to take at best twice the computational time of a trap-free expression, owing to the surface potential equation (SPE) needing to be evaluated at four different locations in contrast to two in the trap-free case. Quantum-mechanical capability was added for determination of inversion charge density, but not accumulated charge density. The results obtained from the triangular well approximation were modified slightly to include reduction in surface field owing to interface states.

Effective mobility was investigated and an analytical correction proposed for a bulk MOS-FET with negligibly low interface state density. The analytical correction is capable of factoring-in the drain bias discrepancy often existing between the capacitance and current measurements. However, it requires that doping level and oxide capacitance density must be known, although the analytical correction is not overly sensitive to the former. A limitation occurs for high values of interface state density $D_{it}$, hindering mobility estimation techniques.

Indirect methods of extracting $D_{it}$ were also discussed by comparing theoretical results with experimental results. The motivation to pursue such methods came from the desire to analyze devices without a reliable bulk contact. While not as precise as direct methods such as charge pumping and the conductance method, the presented indirect methods are useful alternatives because they depend on drain current. Optimization using the ratio of transconductance to drain current below peak transconductance was found to be an excellent means of extracting $D_{it}$ since it converted measured drain current into a “mobility-less” quantity. While the method is prone to measurement noise, it is capable of resolving important parameters such as flatband voltage and surface doping along with $D_{it}$ parameters. Only two parameters (those corresponding to an exponential) were needed to characterize $D_{it}$ because experimental drain current usually corresponds to band bending over a very small region of the bandgap. The technique was applied to Silicon MOSFETs with equivalent oxide thicknesses below 5 nm.
Finally, GaN-based MISFETs were considered using the knowledge built-up over the previous chapters. The data from a GaN-based MISFET suffering from severe gate voltage stretchout were analyzed, and the results provided the conclusion that interface state capacitance density was around two orders of magnitude greater than the oxide capacitance density. The extracted mobility had a very low value and is possibly a consequence of Coulomb scattering caused by the high interface state density (high bulk scattering could also contribute).

7.1 Suggested work

The compact model of Chapter 3 could be further improved. It is possible that a different form of equation (3.19) could be realized that minimizes error shown in Figure 3.5 for arbitrary $V_{DS}$ values. Some of this could be related to determination of precise analytical approximations to interface trapped charge density that are compatible with the U-like $D_{it}$ assumption. When considering interface states, the quantum-mechanical description of inversion charge density could need the value of $\eta$ tailored. This would be best accomplished by thorough application of the conductance method to evaluate interface state density up to the minority carrier band edge. Then, with the $D_{it}$ distribution known, theoretical comparisons to measured capacitance-voltage could be made until the correct value of $\eta$ is found.

If any analytical approximation for $I_{ds}$ could be found in weak through moderate inversion then Method 2 of Chapter 5 could possibly allow for direct extraction of the $D_{it}$ parameters. Moderate inversion roughly corresponds to the near-threshold region of $I_{ds}$ versus gate voltage and is mathematically the most challenging region to model analytically even in the absence of interface states. In this author’s opinion developing the necessary trap-affected approximation is likely impossible unless certain undesirable approximations are made to the expression for $Q_{it}'$. The limitation is imposed by interface state capacitance density $C_{it}'$ not being well-
approximated as $qD_{it}$ near the upper limit of integration used in evaluation of $Q_{it}'$ (Figure 2.14). For this reason optimization was resorted to for parameter extraction with Method 2.

The measured GaN MISFET behavior is indicative of very high interface state density. The addition of a bulk contact would likely add more information as then alternative characterization techniques could be applied. The hysteresis observed after sweeping gate bias makes parameter extraction complicated and likely accounted for the small drain-bias dependence of extracted parameters. For this reason pulsed-IV would be a good candidate. One could choose to hold the drain bias at a desired $V_{DS}$ and the gate bias at some constant value near zero. Then, the gate bias could be pulsed to a desired value, and if done quickly enough could possibly eliminate most of the effects of interface states. Given that interface states respond very quickly in strong-inversion there is uncertainty if this could result in any new information.
REFERENCES


