SURFACE MOBILITY NEAR THRESHOLD AND OTHER PARAMETERS
OF INSULATED GATE FIELD EFFECT TRANSISTORS

BY
Alfred P. Gnadinger

Dipl. El. Ing. ETH Swiss Federal Institute of Technology, Zurich, 1965
M.S.E.E. University of Kansas, Lawrence, 1968

Submitted to the Department of Electrical Engineering and the Faculty of the Graduate School of the University of Kansas, Lawrence, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dissertation Committee:

[Signatures]

Chairman

May, 1970
TO MY WIFE AND OUR PARENTS
ABSTRACT

The mobility of the mobile carriers in the inversion layer of an Insulated Gate Field Effect Transistor (IGFET) has been investigated with particular reference to the threshold or transition region. The surface mobility was deduced from measured channel conductance versus gate voltage curves. The relationship between gate voltage and surface potential was established experimentally using a quasi-static thermal equilibrium C-V curve measured on an MIS capacitor. Influences of interface states are automatically included by using this technique. Experimental hole mobilities obtained on double insulator structures (Al-Al$_2$O$_3$-SiO$_2$-Si) are presented. It was found that the surface mobility of holes near the transition region is of the order of the bulk mobility and decreases with increasing surface potential. A reasonable match between the experimental values and the theory has been obtained.

Some of the IGFET parameters such as interface states, flatband voltage, threshold voltage and surface doping density are critically discussed. New techniques are introduced for measuring interface state density and flatband voltage in order to obtain the required accuracy for investigation of the threshold region of IGFETs.

The carrier distribution in the inversion layer of an MIS structure at room temperature was calculated assuming a quantization of the allowed energy levels at the surface and a linear electrostatic potential. It was found that for strong inversion the carrier distribution deviates considerably from the one found using classical statistics but approaches the classical limit for weak inversion when many electric subbands are occupied. A new definition of the channel thickness is introduced based on the total
charge in the channel and compared to other definitions. Channel thicknesses so defined range from 30 to 400 Å for practical devices depending on surface potential. The total charge in the channel deviated little from the one found by using classical statistics. The channel thickness decreases with increasing surface potential and increases from source to drain if the IGFET is operated in the saturation mode, contrary to the behavior of the conventional channel thickness—the intrinsic distance—which increases with increasing surface potential and decreases from source to drain if the IGFET is operated in the saturation mode.
ACKNOWLEDGEMENT

I wish to express my sincere thanks to my adviser, Professor Harry E. Talley of the Department of Electrical Engineering who provided guidance and continuous encouragement through the course of this thesis.

The financial support by the Graduate School of the University of Kansas and the computation time provided by the University of Kansas Computation Center are gratefully acknowledged.

Appreciation is expressed to the members of my graduate committee, Professor Don G. Daugherty and especially Professor Leroy Colquitt, Jr., of the Department of Physics, who offered helpful suggestions during the preparation of this thesis.

I would like to thank especially Bell Telephone Laboratories, Allentown, Pa., where most of the experimental part of this thesis was completed during two summer employments in 1968 and 1969, and particularly Drs. E. F. Labuda, H. E. Nigh and J. Godfrey. Bell Telephone Laboratories also furnished all the experimental devices used in this work.

Last but by no means least, I wish to express my deepest gratitude to my wife, Susanna. Besides presenting me with a handsome son shortly before completion of this thesis, she has helped me greatly by her encouragement and understanding.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>IGFET STRUCTURE AND PRINCIPLES OF OPERATION</td>
<td>8</td>
</tr>
<tr>
<td>II.1</td>
<td>Principles of Operation</td>
<td>8</td>
</tr>
<tr>
<td>II.2</td>
<td>Description of the Experimental IGFET</td>
<td>12</td>
</tr>
<tr>
<td>III</td>
<td>BASIC IGFET PARAMETERS</td>
<td>18</td>
</tr>
<tr>
<td>III.1</td>
<td>Surface States</td>
<td>18</td>
</tr>
<tr>
<td>III.1.1</td>
<td>Conductance Method</td>
<td>23</td>
</tr>
<tr>
<td>III.1.2</td>
<td>Capacitance Method</td>
<td>24</td>
</tr>
<tr>
<td>III.2</td>
<td>Flatband Voltage</td>
<td>34</td>
</tr>
<tr>
<td>III.3</td>
<td>Threshold Voltage</td>
<td>47</td>
</tr>
<tr>
<td>III.4</td>
<td>Surface Doping Density</td>
<td>50</td>
</tr>
<tr>
<td>III.4.1</td>
<td>Surface Doping Density Determined from High Frequency C-V Curve</td>
<td>51</td>
</tr>
<tr>
<td>III.4.2</td>
<td>Alternative Methods for Determining Surface Doping Density</td>
<td>55</td>
</tr>
<tr>
<td>IV</td>
<td>IGFET INVERSION LAYER</td>
<td>61</td>
</tr>
<tr>
<td>IV.1</td>
<td>Carrier Distribution: Quantum Mechanical Model</td>
<td>63</td>
</tr>
<tr>
<td>IV.2</td>
<td>Carrier Distribution: Classical Model</td>
<td>77</td>
</tr>
<tr>
<td>IV.3</td>
<td>Channel Thickness</td>
<td>79</td>
</tr>
<tr>
<td>V</td>
<td>SURFACE MOBILITY NEAR THE TRANSITION REGION</td>
<td>92</td>
</tr>
<tr>
<td>V.1</td>
<td>Theories of Surface Mobility</td>
<td>92</td>
</tr>
<tr>
<td>V.2</td>
<td>Experimental Determination of Surface Mobility</td>
<td>103</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>V.3 Experimental Results</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>V.4 Conclusions</td>
<td>114</td>
<td></td>
</tr>
<tr>
<td>VI SUMMARY</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>APPENDIX I</td>
<td>128</td>
<td></td>
</tr>
</tbody>
</table>
## LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Area</td>
</tr>
<tr>
<td>A</td>
<td>Area of unit cell in k-space</td>
</tr>
<tr>
<td>C</td>
<td>Capacitance per unit area</td>
</tr>
<tr>
<td>C_i</td>
<td>Insulator capacitance per unit area</td>
</tr>
<tr>
<td>C_{max}</td>
<td>Maximum high frequency MIS capacitance</td>
</tr>
<tr>
<td>C_{min}</td>
<td>Minimum high frequency capacitance</td>
</tr>
<tr>
<td>C_{sc}</td>
<td>Theoretical incremental low frequency semiconductor surface capacitance per unit area</td>
</tr>
<tr>
<td>D</td>
<td>Effective Density of States</td>
</tr>
<tr>
<td>D_1</td>
<td>Effective Density of States in xy-plane</td>
</tr>
<tr>
<td>D_2</td>
<td>Effective Density of States in z-direction</td>
</tr>
<tr>
<td>E</td>
<td>Energy</td>
</tr>
<tr>
<td>E</td>
<td>Surface energy</td>
</tr>
<tr>
<td>E_F</td>
<td>Fermi energy</td>
</tr>
<tr>
<td>E_g</td>
<td>Band gap energy</td>
</tr>
<tr>
<td>E_i</td>
<td>i^{th} quantized energy level</td>
</tr>
<tr>
<td>E_m</td>
<td>Highest energy level</td>
</tr>
<tr>
<td>E_1</td>
<td>First energy level</td>
</tr>
<tr>
<td>ΔE</td>
<td>Energy difference</td>
</tr>
<tr>
<td>E_p</td>
<td>Pinch-off field</td>
</tr>
<tr>
<td>E_x</td>
<td>Electric field in x-direction</td>
</tr>
<tr>
<td>E_{xs}</td>
<td>Electric field in x-direction at the surface</td>
</tr>
<tr>
<td>F(u,u_p)</td>
<td>Kingston - Neustadter F-function</td>
</tr>
<tr>
<td>F(u,ξ,u_F)</td>
<td>Nonequilibrium Kingston - Neustadter F-function</td>
</tr>
<tr>
<td>F_o</td>
<td>Zero order Fermi - Dirac Integral</td>
</tr>
<tr>
<td>F_{01}</td>
<td>F_0 corresponding to first energy level</td>
</tr>
<tr>
<td>F_{0max}</td>
<td>F_0 corresponding to highest energy level</td>
</tr>
<tr>
<td>H</td>
<td>Function defined in equation 4.14</td>
</tr>
<tr>
<td>I_D</td>
<td>Drain current</td>
</tr>
<tr>
<td>I_{Dsat}</td>
<td>Saturation drain current</td>
</tr>
<tr>
<td>I_S</td>
<td>Reverse bias saturation current of pn-junction</td>
</tr>
<tr>
<td>K_o</td>
<td>Defined in equation 5.7</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>L</td>
<td>Channel length</td>
</tr>
<tr>
<td>$L_D$</td>
<td>Intrinsic Debye length</td>
</tr>
<tr>
<td>N</td>
<td>Total number of electrons in the inversion layer</td>
</tr>
<tr>
<td>$N_{A}$</td>
<td>Acceptor doping density</td>
</tr>
<tr>
<td>$N_{D}$</td>
<td>Donor doping density</td>
</tr>
<tr>
<td>$N_{i}$</td>
<td>Number of electrons occupying the $i^{th}$ energy level</td>
</tr>
<tr>
<td>$N_{SS}$</td>
<td>Interface state density</td>
</tr>
<tr>
<td>$Q_D$</td>
<td>Depletion charge per unit area</td>
</tr>
<tr>
<td>$Q_P$</td>
<td>Mobile charge in the inversion layer per unit area</td>
</tr>
<tr>
<td>$Q^*_P$</td>
<td>Mobile charge limit for determining $V_T$</td>
</tr>
<tr>
<td>$Q_S$</td>
<td>Total induced charge in the semiconductor per unit area</td>
</tr>
<tr>
<td>$Q^*_{SS}$</td>
<td>Fixed charge in the insulator at the semiconductor - insulator interface per unit area</td>
</tr>
<tr>
<td>$Q^*_{t}$</td>
<td>Charge trapped in interface states</td>
</tr>
<tr>
<td>$\Delta Q$</td>
<td>Charge difference</td>
</tr>
<tr>
<td>R</td>
<td>Resistance</td>
</tr>
<tr>
<td>$T$</td>
<td>Absolute temperature</td>
</tr>
<tr>
<td>$T'$</td>
<td>Time period of ac-signal (equation A-11)</td>
</tr>
<tr>
<td>$U(x)$</td>
<td>Electron potential</td>
</tr>
<tr>
<td>$V$</td>
<td>Voltage</td>
</tr>
<tr>
<td>$V_{acc}$</td>
<td>Voltage corresponding to strong accumulation</td>
</tr>
<tr>
<td>$V_D$</td>
<td>Drain to source voltage</td>
</tr>
<tr>
<td>$V_{Dsat}$</td>
<td>Saturation drain voltage</td>
</tr>
<tr>
<td>$V_{DO}$</td>
<td>Amplitude of ac drain voltage</td>
</tr>
<tr>
<td>$V_{FB}$</td>
<td>Flatband voltage</td>
</tr>
<tr>
<td>$V_G$</td>
<td>Gate to source voltage</td>
</tr>
<tr>
<td>$V^{'}_G$</td>
<td>$V_G - V_{FB}$, effective gate voltage</td>
</tr>
<tr>
<td>$V_O$</td>
<td>Flatband voltage without interface states, defined in equation 3.11</td>
</tr>
<tr>
<td>$V_T$</td>
<td>Threshold voltage</td>
</tr>
<tr>
<td>$V^*_T$</td>
<td>Intercept of linear region of $Q_P(V_G)$ with voltage axis for ideal structure</td>
</tr>
<tr>
<td>$\Delta V$</td>
<td>Voltage shift</td>
</tr>
<tr>
<td>$\Delta V_S$</td>
<td>Change in voltage drop over semiconductor</td>
</tr>
<tr>
<td>$W$</td>
<td>Channel width</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>a</td>
<td>Slope of voltage ramp</td>
</tr>
<tr>
<td>a</td>
<td>Defined in equation 4.29</td>
</tr>
<tr>
<td>c_i</td>
<td>Normalization constant</td>
</tr>
<tr>
<td>d</td>
<td>Insulator thickness</td>
</tr>
<tr>
<td>f(E)</td>
<td>Fermi-Dirac distribution</td>
</tr>
<tr>
<td>f(t)</td>
<td>General function of time</td>
</tr>
<tr>
<td>f_{eff}</td>
<td>Effective value of f(t)</td>
</tr>
<tr>
<td>( g_C )</td>
<td>Small signal conductance of the channel proper without external conductances</td>
</tr>
<tr>
<td>( g_{CO} )</td>
<td>Zero drain voltage conductance of the channel proper</td>
</tr>
<tr>
<td>( g_{C}^m )</td>
<td>Measured value of ( g_C )</td>
</tr>
<tr>
<td>( g_{CO}^m )</td>
<td>Measured value of ( g_{CO} )</td>
</tr>
<tr>
<td>( g_D )</td>
<td>Small signal channel conductance</td>
</tr>
<tr>
<td>( g_{DO} )</td>
<td>Zero drain voltage channel conductance</td>
</tr>
<tr>
<td>( g_D^m )</td>
<td>Measured value of ( g_D )</td>
</tr>
<tr>
<td>( g_J )</td>
<td>Conductance of the drain (or source) to substrate junction</td>
</tr>
<tr>
<td>( g_{JO} )</td>
<td>Zero drain (or source) voltage value of ( g_J )</td>
</tr>
<tr>
<td>( g_{J}^m )</td>
<td>Measured value of ( g_J )</td>
</tr>
<tr>
<td>( g^m )</td>
<td>Measured conductance</td>
</tr>
<tr>
<td>( g_1 )</td>
<td>Upper conductance limit in determining flatband voltage</td>
</tr>
<tr>
<td>( g_2 )</td>
<td>Lower conductance limit in determining flatband voltage</td>
</tr>
<tr>
<td>( \hbar )</td>
<td>( = h/2\pi ), Planck's constant</td>
</tr>
<tr>
<td>i</td>
<td>Current</td>
</tr>
<tr>
<td>i_{eff}</td>
<td>Effective value of current</td>
</tr>
<tr>
<td>k</td>
<td>Wave vector</td>
</tr>
<tr>
<td>k</td>
<td>Boltzmann's constant</td>
</tr>
<tr>
<td>m_0</td>
<td>Rest mass of electron</td>
</tr>
<tr>
<td>m^*</td>
<td>Effective mass</td>
</tr>
<tr>
<td>m_\sigma</td>
<td>Conductivity effective mass</td>
</tr>
<tr>
<td>n</td>
<td>Electron concentration</td>
</tr>
<tr>
<td>n_i</td>
<td>Intrinsic carrier concentration</td>
</tr>
<tr>
<td>n_0</td>
<td>Equilibrium electron concentration</td>
</tr>
<tr>
<td>p</td>
<td>Hole concentration</td>
</tr>
<tr>
<td>p_b</td>
<td>Equilibrium hole concentration</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>Δp</td>
<td>Difference in hole concentration defined in eq. 5.6</td>
</tr>
<tr>
<td>q</td>
<td>Electronic charge</td>
</tr>
<tr>
<td>r</td>
<td>Hall ratio</td>
</tr>
<tr>
<td>s_i</td>
<td>i\textsuperscript{th} solution of equation 4.3</td>
</tr>
<tr>
<td>u</td>
<td>Electrostatic potential in kT/q units</td>
</tr>
<tr>
<td>u_D</td>
<td>Drain to source voltage in kT/q units</td>
</tr>
<tr>
<td>u_G</td>
<td>Gate to source voltage in kT/q units</td>
</tr>
<tr>
<td>u_0</td>
<td>Amplitude of ac signal in kT/q units</td>
</tr>
<tr>
<td>u_F</td>
<td>Fermi potential in kT/q units</td>
</tr>
<tr>
<td>u_s</td>
<td>Surface potential in kT/q units</td>
</tr>
<tr>
<td>\dot{u}_s</td>
<td>Sign of u_s</td>
</tr>
<tr>
<td>\vec{V}</td>
<td>Velocity</td>
</tr>
<tr>
<td>v_D</td>
<td>AC drain voltage</td>
</tr>
<tr>
<td>v_{eff}</td>
<td>Effective value of ac voltage</td>
</tr>
<tr>
<td>x</td>
<td>Distance normal to semiconductor surface</td>
</tr>
<tr>
<td>x_i</td>
<td>Intrinsic distance</td>
</tr>
<tr>
<td>y</td>
<td>Distance along the channel</td>
</tr>
<tr>
<td>y_p</td>
<td>Pinch-off distance in the channel</td>
</tr>
<tr>
<td>z</td>
<td>Distance normal to x and y</td>
</tr>
<tr>
<td>Γ</td>
<td>Number of states</td>
</tr>
<tr>
<td>\Lambda</td>
<td>Surface potential difference defined in equation 3.6</td>
</tr>
<tr>
<td>\xi</td>
<td>Electric field</td>
</tr>
<tr>
<td>\phi_{MS}'</td>
<td>Metal-semiconductor workfunction difference</td>
</tr>
<tr>
<td>\phi_{ii}</td>
<td>Additional contact potential in double insulator structures</td>
</tr>
<tr>
<td>\Psi</td>
<td>Wave function</td>
</tr>
<tr>
<td>α</td>
<td>Parameter defined in equation 3.25</td>
</tr>
<tr>
<td>α</td>
<td>Parameter defined in equation 5.4</td>
</tr>
<tr>
<td>α</td>
<td>Angle defined in Fig.A-7</td>
</tr>
<tr>
<td>β</td>
<td>Angle defined in Fig.A-7</td>
</tr>
<tr>
<td>δ_c</td>
<td>Channel thickness</td>
</tr>
<tr>
<td>\epsilon_s</td>
<td>Dielectric constant of the semiconductor</td>
</tr>
<tr>
<td>η</td>
<td>Variable defined in equation 4.18</td>
</tr>
<tr>
<td>\lambda_p</td>
<td>Defined in equation 5.8</td>
</tr>
</tbody>
</table>
\( \mu \) Mobility
\( \mu_b \) Bulk mobility
\( \mu_{\text{eff}} \) Effective mobility
\( \mu_H \) Hall mobility
\( \mu_p \) Hole mobility
\( \nu \) Integration variable (eq. 5.5)
\( \tau \) Relaxation time
\( \tau_b \) Bulk relaxation time
\( \xi \) Variable defined in equation A-3
\( \zeta \) Variable defined in equation 4.18
Chapter I

INTRODUCTION

The Insulated Gate Field Effect Transistor, called IGFET for convenience in this thesis, was perhaps the first semiconductor device to be conceived. In fact, the original description of the principles of the IGFET dates back to the early 1930's when Lilienfeld\(^1\) in the United States and Heil\(^2\) in England proposed using the surface field effect to achieve a solid state amplifier. Unfortunately, because of the limited knowledge of the physics of surfaces and thin films which existed at that time, it did not work. Little advance in the perfection of these devices was made, and interest in them waned with the development of the vacuum tube.

The idea of the IGFET was caught at again in the 1940's when a Bell Laboratories group\(^3\) was actively investigating semiconductors for possible use as microwave detectors. The IGFET again didn't work and the more or less accidental discovery of the bipolar transistor gave a specific direction to solid state device research and development for more than a decade.

The main reasons why the realization of the IGFET was so difficult and took such a long time were:

a) A way had to be found to reduce the high surface state density encountered at the surface of a semiconductor by at least two orders of magnitude. It was shown, mainly by Bardeen\(^4\) and Shockley and Pearson\(^5\), that the mobile charge density induced by an electric field normal to the surface of a semiconductor was very small. Roughly ninety percent of the charges moved by the electric field were unavailable for conduction because they were trapped in surface states.
b) A dielectric had to be found which could be made thin enough in order to achieve high electric fields at the semiconductor surface but which would not break down first.

Fortunately, both these problems were solved simultaneously with the advent of planar technology, which was actually developed for bipolar transistors. The essentials of this technology are the use of a thermally grown silicon dioxide layer which protects the silicon surface and acts at the same time as a diffusion mask. Such a thermally grown oxide layer is an insulator with a high dielectric breakdown. It can be grown as thin as a few hundred Angstroms and this thin layer is well controllable by monitoring the oxidation time. At the same time, the interface state density at the silicon dioxide-silicon interface can be very low, of the order of $10^{11} \text{ cm}^{-2}$, if the oxides are carefully grown and the devices are properly annealed. Planar technology should, therefore, be especially suited for fabricating IGFETs. In 1960 Kahng and Atalla\(^5\) built the first workable IGFET units using such thermally oxidized silicon structure.

The IGFET characteristics have been subsequently studied by Ihantola\(^6\), Ihantola and Moll\(^7\), Sah\(^8\), Hofstein and Heiman\(^9\). In Chapter II of this thesis a short phenomenological description of the IGFET characteristics will be given. A review of the physics, technology and application of IGFETs has been given in a book by Wallmark and Johnson\(^10\).

The IGFETs of the early 1960's were far from perfect. They were highly unstable; i.e., their device characteristics drifted in time under applied temperature and bias stress. Their transconductance was quite low. Overall, IGFETs were inferior to the well perfected bipolar transistors in most respects. In the last years, however, a great effort has been made to improve the device performance. Three directions of research
are perhaps most significant. The first of these goals was finding a technology which allowed fabrication of stable devices. The second one was an effort to improve the theoretical understanding of the MIS system, and the third goal was to reduce the surface state density at the SiO₂-Si interface. We will now look at these three problems more closely.

1) Stability of the IGFET:

It was first demonstrated by Snow et al.\(^{(11)}\) that alkali ions such as sodium in the thermally grown SiO₂ films are mainly responsible for the instability of the IGFET. Sodium ions are mobile enough to be able to migrate through the oxide layer if an electric field is applied. Later Hofstein\(^{(12)}\) showed that hydrogen ions (H\(^+\)) can produce a similar drift effect. The ion drift has now been brought under control by:

a) modification of the insulator, e.g., phosphorus doping on the Si-SiO₂ system or use of an insulator such as silicon nitride which is immune to ion drift and
b) clean fabrication so as to avoid contamination by sodium and other impurities.

Other mechanisms such as surface charge instability, slow trapping instability\(^{(13)}\), etc., can also degrade the device performance. However, the improved oxidation techniques now available have greatly reduced their influence.

2) Theoretical understanding of the semiconductor-insulator interface:

Theoretical understanding of interfaces has greatly improved over the past ten years, especially the understanding of the silicon-silicon dioxide interface. This theoretical knowledge was the basis for many improvements in device processing such as oxidizing and annealing techniques.
A theoretical understanding of the failure mechanisms associated with the interface helped greatly to improve the reliability and stability of the devices. It can well be said that no other interface is as well understood as the Si-SiO$_2$ system.

3) Reduction of surface states:

The early IGFETs had a low transconductance because the surface state density at the SiO$_2$-Si interface was too large. Interface states are also believed to be responsible for the noise in IGFETs. Kahng and Atalla's (5) devices had surface state densities of roughly $10^{12}$ states/cm$^2$. In today's structures, however, surface state densities of less than $10^{10}$ states/cm$^2$ can be achieved routinely by the use of improved annealing techniques. The noise behavior and the transconductance of the devices is considerably improved by the reduction in surface states. Lower surface state densities also improve the stability of the devices by reducing the slow trapping instability (13).

Another research goal was to increase the cutoff frequency of IGFETs and make them applicable for operation in the high frequency domain. This was mainly done by reducing the channel length and using a semiconductor with a high surface mobility such as GaAs. It should be mentioned, however, that the IGFET will probably never be able to operate at very high frequencies (gigahertz range), since other devices such as the Schottky Barrier Gate Field Effect Transistor (14) are better suited and will take its place.

Although IGFETs have been made on various semiconductors such as Ge (15), Si (5) and GaAs (16) and have used various insulators such as SiO$_2$ (5), Si$_3$N$_4$ (17) and Al$_2$O$_3$ (18), by far the most important and most often used system is the Si-SiO$_2$ combination.
In summary of the foregoing remarks, it can well be said that Silicon Insulated Gate Field Effect Transistors are now highly perfected devices and excellent device performance is achieved routinely using standardized processes.

The goal of this thesis is to contribute to the understanding of some features of IGFETs which have received little attention in the past—either because they are of a more academic interest (Chapter IV)—or because until now there was no need for understanding them in order to be able to fabricate useful devices (Chapter V). Experimental results were obtained on Silicon IGFETs with an Al₂O₃-SiO₂ sandwich as gate insulator. These particular structures, which are described in detail in Chapter II, are not widely used for commercial devices. However, they have the following advantages over the more commonly used metal-SiO₂-silicon system:

a) Al₂O₃ controls the ion drift in the gate insulator much in the same way as the commonly used silicon nitride.
b) For some applications a low threshold voltage (~1V) is desired. Al₂O₃ introduces an additional contact potential, called \( \phi \) in the literature, which reduces the threshold voltage.
c) The advantages of the well controllable SiO₂-Si interface, especially the low surface state density, are preserved.

IGFETs with an Al₂O₃-SiO₂ insulator were first fabricated and described by H. E. Nigh et al. (19) at Bell Telephone Laboratories.

All experimental results reported in this thesis were obtained on enhancement mode p-channel devices. Enhancement mode n-channel IGFETs are still difficult to fabricate because of the magnitude of the positive oxide charge which is unavoidably deposited (see Chapter III). For consistency with the experiments, wherever possible the theories developed in this thesis are given for p-channel devices as well. A detailed
description of the experimental IGFETs and of the processes used for their fabrication is given in Chapter II.

Chapter III deals with some IGFET parameters such as surface states, flatband voltage, threshold voltage and surface doping density whose definitions and methods of measurement are critically discussed. Where necessary new measurement techniques have been introduced.

Chapter IV is devoted to the nature of the conducting channel of the IGFET. Particular attention is paid to the carrier distribution within the inversion layer, which is calculated using a classical and a quantum mechanical model. A new and more realistic definition of the thickness of the channel is introduced and applied to the classical and the quantum-mechanical description of the channel. As a consequence of this channel thickness definition, a new concept of the shape of the channel in the pinched-off mode of the IGFET emerges.

A feature of the IGFET which has received little attention in the past is the transition region between depletion and inversion (region near threshold)\(^{20}\). One reason is that the devices are usually operated at gate voltages well above threshold because the transconductance increases with gate voltage and the effect of threshold voltage variation caused by instabilities are diminished. However, the gate voltage range near threshold will be of increased interest in the future because lower supply voltages tend to be used (e.g., a 3V supply voltage for a device with 1V threshold voltage) as the device technology is improved. Besides reducing the power dissipation, a low supply voltage makes the IGFET more compatible with bipolar transistors. Chapter V is devoted to this transition region. The main parameter investigated is the surface mobility, which is still not well understood. Various theories\(^{21}\)-(\(^{23}\)) have been
offered, but a convincing match between experimental values and those theories is still lacking. Especially, no attempt has been made so far to characterize the surface mobility in the transition region. A new approach is chosen in this work to find experimental values of surface mobility in or near this transition region. These values are then compared to the most reasonable theory.

Finally, Chapter VI gives a summary of the topics covered in this thesis and the results obtained. Suggestions for future work are given at the end of this chapter.
Although various kinds of IGFET structures can be and have been fabricated the most common one is of the rectangular type as shown schematically in Fig. 2.1. The coordinate system introduced in Fig. 2.1 will be used throughout this thesis. As mentioned in the introduction we will only consider p-channel (n-type substrate) devices. The following is a short description of the principles of operation and the device characteristics of an IGFET.

II.1 Principles of Operation

As shown in Fig. 2.1 the IGFET consists of an n-type silicon substrate into which two p+-regions, the source and the drain, are diffused. The region between the source and the drain is under the influence of a metal field plate or gate. If a large enough negative voltage is applied to the gate, the surface of the underlying silicon can be inverted and a conductive p-channel is formed connecting source and drain. The conductivity of this channel can then be modulated by varying the gate voltage. The drain voltage (voltage between drain and source) is negative in the normal mode of operation of a p-channel device. The cases of small and large drain voltages are considered separately. For small drain voltages the channel induced between drain and source behaves essentially like a resistor. Its resistance is (24):

\[ R = \frac{L}{W \mu_p Q_p} \]  

(2.1)
Fig. 2.1 Schematic diagram of an insulated-gate field-effect transistor. The important parameters are the channel length \((L)\), the channel width \((W)\), and the insulator thickness \((d)\). (After Kahng and Atalla, Ref. 5)

Fig. 2.2 Current-voltage characteristics of a silicon p-channel IGFET with \(W/L=25\), \(d=0.08\mu\), \(N_D=3 \times 10^{16}\text{cm}^{-3}\), \(V_{FB}=+2\text{V}\).
where \( L \) and \( W \) are the length and width of the channel, respectively, \( \mu_p \) the mobility and \( Q_p \) the charge per unit surface area of holes in the inversion layer. The resistance \( R \) depends on the magnitude of \( Q_p \) as is obvious from (2.1). \( Q_p \) itself depends on the gate voltage and the properties of the semiconductor.

As the magnitude of the drain voltage is now increased eventually \( Q_p \) becomes a function of position in the channel and (2.1) is not valid any more. The potential difference between the gate and the inversion layer decreases if we average it over the channel length. As a result the average \( Q_p \) will also decrease and the resistance of the channel increases. Thus the drain current versus drain voltage curve will begin to bend downward from the initial resistor line. This is evident in the characteristics of a p-channel IGFET shown in Fig. 2.2 for various gate voltages.

As the drain voltage is further increased the voltage drop across the oxide near the drain eventually falls below a level required to maintain an inversion layer. The drain voltage at which this happens will be denoted by \( V_{Dsat} \). The concentration of holes near the drain falls to a level much smaller than in the region near the source. The surface near the drain is essentially depleted of mobile carriers. If the magnitude of the drain voltage is increased still further the IGFET operates in the saturation mode. The potential at the end of the inversion layer remains constant at the value \( V_{Dsat} \) and the drain voltage in excess of \( V_{Dsat} \) will drop over the depletion region near the drain. The magnitude of the

* All mathematical symbols used in this thesis are defined in the "List of Symbols."
current flowing in the surface inversion layer is essentially determined by the voltage drop over the inversion layer. Since this voltage drop remains practically constant for the saturation mode the current will not change significantly with drain voltage and will remain at the value \( I_{\text{Dsat}} \) corresponding to \( V_{\text{Dsat}} \) as is evident from Fig. 2.2.

As the gate voltage is increased the saturation drain voltage \( V_{\text{Dsat}} \) also increases since a larger voltage is now required to pinchoff a channel with a larger conductance. The saturation current \( I_{\text{Dsat}} \) also increases with increasing gate voltage, as is evident from Fig. 2.2.

Recapitulating this description, we can distinguish two regions of operation of an IGFET:

a) For small drain voltages the current-voltage characteristics are nearly ohmic or linear. This is the \textit{linear region}.

b) For large drain voltages the current is nearly independent of drain voltage. This is the \textit{saturation region}.

For drain voltages \(|V_D| > |V_{\text{Dsat}}|\) the drain current does not, however, saturate completely but increases slightly with increasing \(|V_D|\). Recently several attempts have been made to explain this behavior theoretically. Reddi and Sah\(^{(25)}\) showed that the reduction of the channel length due to a widening of the depletion region at the drain for \(|V_D| > |V_{\text{Dsat}}|\) can account for a finite saturation resistance. Pao and Sah\(^{(26)}\) included the diffusion current in the analysis and obtained an analytical expression for both the linear and the saturation region of an IGFET with a finite resistance in the saturation region. Grosvalet et al.\(^{(27)}\) explain the saturation behavior of the IGFET by assuming a limited velocity of the mobile carriers in the drain region of the
IGFET. Very recently Schröder and Muller\(^{(28)}\) obtained good match between theoretical and experimental saturation drain currents. Making certain approximations, they calculated the quite complicated electrostatic potential distribution in the space charge region of the IGFET by solving the 2-dimensional Poisson's equation numerically. Similar approaches were taken by Geurst\(^{(29)}\) and Wright\(^{(30)}\). Summarizing these theories it can well be said that the "true" theory will be a mixture of these effects or approaches. Depending on the structural parameters such as doping density, oxide thickness, channel length, etc., one or the other will dominate and determine the behavior of the IGFET in saturation.

**II.2 Description of the Experimental IGFETs.**

We will now describe the actual structures used as experimental devices in this thesis and explain the various processing steps necessary for their fabrication.

All experimental IGFETs used in this work are of the rectangular type as shown in Fig. 2.1. The various processing steps for their fabrication are graphically displayed in Fig. 2.3 and will now be discussed in detail.

The IGFETs are made on standard 1\(\frac{3}{4}\)" diameter n-type silicon wafers with a starting resistivity of 3 - 7 ohm-cm corresponding to a doping density of \(7 \times 10^{14} \text{ cm}^{-3}\) to \(1.5 \times 10^{15} \text{ cm}^{-3}\). After cleaning, the wafers undergo an initial oxidation (steam oxide) for 50 min. at 1100°C. Openings for the drain and source islands are then cut into the oxide using conventional photoresist techniques and boron is
Fig. 2.3 Processing steps in the fabrication of $\text{Al}_2\text{O}_3$-$\text{SiO}_2$ double insulator IGFETs.
diffused into these drain and source windows resulting in heavily
doped p⁺-islands. A protective SiO₂ layer is then grown over the whole
slice in a steam ambient (950°C for 150 min.). Next the gate opening
is cut into this oxide and a special SiO₂ layer grown (1100°C, wet at-
mosphere). This oxide determines to a great extent the characteristics
of the IGFET. Therefore, extreme care is taken to reduce the contami-
nation of alkali ions (sodium) and hydrogen ions (see Chapter I). For
example, no methanol is used in all the cleaning steps since it has
been found that methanol causes contamination by H⁺ ions (protons).
The oxidation time for the gate oxide depends on the desired SiO₂ thick-
ness, being 18 min. for the standard thickness of 1000 Å. To reduce
the interface state density and stabilize the gate oxide the slices
are annealed subsequently in a nitrogen atmosphere. As the next step
Al₂O₃ is deposited on the slices using a deposition technique developed
by Tung and Caffrey(31),(32) and H. E. Nigh(19). In this method the
Al₂O₃ layer is formed by reacting aluminum chloride with a carbon di-
oxide - hydrogen mixture at elevated temperatures (900°C). The reac-
tion time for 500 Å Al₂O₃ (standard thickness) is 9.5 min. As a mask
for subsequent etchings an SiO₂ layer (3000 Å) is then deposited on the
slice by oxidation of silane.(33)

Using again standard photoresist techniques the openings for
the drain and source contacts are cut into this SiO₂ layer. After re-
moval of the photoresist the SiO₂ layer is used as a mask to permit
etching through the underlying Al₂O₃ layer using hot phosphoric acid.
The Al₂O₃ layer in turn serves as a mask for etching through the re-
mainin SiO₂ layer. The procedure is the same as that described by
W. van Gelder and V. E. Hauser(34) for the etching of silicon nitride.
After a special cleaning step the slice is metallized by depositing 200 Å titanium and 5000 Å aluminum by standard vacuum evaporation techniques. In a final photolithographic step the unwanted parts of the metal layer are etched away and the desired metallization pattern is formed.

Fig. 2.4 shows a top view of a finished IGFET. Most slices processed contain two different IGFET geometries in alternating rows (20-40 units per row) and a row of capacitors with the same gate insulator as the IGFET along a center line of the wafer. The first IGFET-type, called FET 3, has a channel length $L$ of 5 $\mu m$ and a channel width $W$ of 50 $\mu m$ (See Fig. 2.1). The second IGFET-type, FET 4, has the same channel length $L$ but a channel width $W$ of 100 $\mu m$. As will be shown later it is advantageous to make the $W/L$ ratio as large as possible. The channel length $L$ of 5 $\mu m$ employed in these structures represents approximately the lower limit obtainable with today's technology. If the channel length $L$ is further decreased, unwanted effects appear such as the punch-through effect (6) or at least a decrease in the saturation resistance (25). The field plates of the capacitors consist of square 10 x 10 mil metal dots ($6.45 \times 10^{-4} \text{ cm}^{-2}$).

Different insulator thicknesses, $\text{SiO}_2$ as well as $\text{Al}_2\text{O}_3$, are used in the experimental structures. The $\text{Al}_2\text{O}_3$ thicknesses range from 500 Å to 2000 Å. Layers thinner than 300 Å are nonuniform and unstable. $\text{Al}_2\text{O}_3$ layers thicker than approximately 3000 Å show a tendency to crack. The standard $\text{Al}_2\text{O}_3$ thickness most often used is 500 Å. The thicknesses of the underlying $\text{SiO}_2$ layer range from 500 Å to 4000 Å. $\text{SiO}_2$ layers thinner than 200 Å show trapping instabilities (tunneling of carriers to the $\text{SiO}_2 - \text{Al}_2\text{O}_3$ interface), whereas layers much thicker than 4000 Å make the oxide capacitance too small for a useful device. The most often used $\text{SiO}_2$ thickness is 1000 Å.
Fig. 2.4 Photomicrographs of finished IGFETs and MIS capacitor.

a) MIS capacitor: 10 x 10 mils square.
   FET 3: W/L = 10, L = 5\(\mu\)m
b) FET 4: W/L = 20, L = 5\(\mu\)m
For all slices a (100) surface orientation is chosen since this gives the lowest possible surface charge and surface state densities. For commercial devices, the same orientation is usually employed.
Chapter III

BASIC IGFET PARAMETERS

Among the many parameters which govern the behavior of IGFETs, the surface states (interface states and fixed surface charge), the flatband voltage, the surface doping density and the threshold voltage will be covered in this chapter. These parameters are of fundamental significance for the operation of an IGFET. Their definitions and the methods of measurement are investigated and critically discussed. Special attention is paid to their usefulness in the transition or threshold region of the IGFET. Where necessary new definitions and new measurement techniques are introduced for that purpose.

No attempt is made to investigate and describe the high art of getting these surface parameters under control, i.e., to obtain certain prescribed values of these parameters. The emphasis is put on the methods of measurement.

III.1 Surface States

As mentioned in Chapter I surface states at the semiconductor surface were the reason the early IGFETs did not work. It is, therefore, of utmost importance to understand their properties theoretically, to get them under control and, hopefully, to reduce them to a low enough level in order to fabricate useful IGFETs. The surface states have been theoretically studied by Tamm(35),(36), Shockley(37) and others(38),(39) and have been shown to exist within the forbidden gap due to the interruption of the periodic lattice structures at the
surface of a crystal. The existence of surface states was first found experimentally by Shockley and Pearson\(^3\) in their surface conductance measurement.

Before MIS technology was developed the surface states were investigated on either "clean" surfaces which were usually obtained by cleaving a crystal in an ultrahigh vacuum, or on "real" surfaces. The term "real" surface means the surface of a polished and etched wafer with a natural oxide layer which can range from 10 Å if the wafer is freshly cleaned to 30 Å if the wafer is several months old. Measurements on clean surfaces confirmed the theoretical prediction that the density of surface states is very high—of the order of the density of surface atoms\(^{40}\), being \(8 \times 10^{14} \text{ cm}^{-2}\) on the (111) plane of Si. For "real" surfaces it was found that the surface state density could be much lower \((10^{12} - 10^{13} \text{ cm}^{-2})\).\(^{41}\) Historically, surface states have been classified into fast and slow states. The fast states exchange charge with the conduction or valence band rapidly and were assumed to lie close to the interface between the semiconductor and the insulator. Slow states, on the other hand, were believed to exist at the interface of the air and the thin (10 Å to 30 Å) insulator and require a longer time for charge exchange.

For the present MIS system with thick insulating layers (≈1000 Å) this classification is not very useful. The "slow" states at the air-insulator interface have now essentially infinite time constants; they cannot exchange charge with the semiconductor. In an experiment they have negligible influence because they are far away from the semiconductor-insulator interface. Even if they do influence the semiconductor, they cannot be distinguished from the fixed oxide charge. The
only states that we are concerned with are the surface states or interface states at the insulator-semiconductor interface. And these states are not necessarily fast surface states either, since at low temperatures, the time constant of these states can be very long\(^{(42),(43)}\).

There is still controversy in the literature as to how best to classify the surface states at the semiconductor-insulator interface. The following terminology is proposed and will be used throughout this thesis:

1) **Interface states** are those surface states which can exchange charge with the semiconductor if the surface potential is changed slowly enough. Their density is denoted by \(N_{ss} (\text{cm}^{-2}\text{ev}^{-1})\). What to understand by "slowly enough" will be explained later.

2) **Surface charge** is the fixed charge, denoted by \(Q_{ss} (\text{cm}^{-2})\), at the semiconductor-insulator interface. Part of \(Q_{ss}\) can actually be a distributed space charge in the insulator. This space charge is treated as if it were located at the interface with its magnitude adjusted so that the influence on the semiconductor is the same.

As an illustration the band structure of an inverted silicon surface (MIS capacitor) and the two types of surface states are shown schematically in Fig. 3.1. Since the interface states can exchange charge with the semiconductor their occupancy is governed essentially by Fermi-Dirac statistics. That is, below the Fermi-level the states are occupied by electrons, above the Fermi-level they are unoccupied. No attempt will be made here to discuss the nature and the properties of the interface states and the surface charge. For good review articles see for example Deal et al.\(^{(44)}\) and Revesz and Zaininger\(^{(45)}\). We will only be concerned with their measurement.

The distinction between interface states and surface charge can become ambiguous if interface states with very long time constants are present. Even at room temperature this situation can occur at SiO\(_2\)-Si.
Fig. 3.1 Energy band diagram of an n-type MIS capacitor in inversion and classification of surface states
structures irradiated by $\gamma$-rays.\(^{(46)}\) In this thesis, however, a clear distinction could always be made between surface charge and interface states since no states with long ($>1$ second) but finite time constants were present in our samples. The term "slowly enough" in the definition of the interface states does, therefore, mean that the surface potential has to change so slowly that all states with finite time constants can follow.

The reason the distributed charge in the insulator is lumped together with the surface charge is (a) no distinction can be made in its measurements and (b) the distinction would be of no importance for the device performance. One should, however, always keep in mind that a numerical value obtained for $Q_{ss}$ does not have to be the actual charge density at the semiconductor interface. It can include influences of uncompensated space charge within the insulator. This is particularly true for the double insulator structures used in this thesis and described in Chapter II. Any free charge on the $\textrm{Al}_2\textrm{O}_3$-$\textrm{SiO}_2$ interface can alter the value of $Q_{ss}$.

The different methods of measuring interface state densities will now be discussed. Methods of measuring surface charge will be treated in Section III.2.

**Measurement of Interface State Density**

Interface states are most conveniently measured on MIS capacitors. In this work, interface states are measured on the 10 x 10 mil square metal dots located on the center line of the IGFET wafer (see Chapter II). The information about interface states can either be
extracted from a measured capacitance versus voltage curve (capacitance method) or from a conductance versus voltage curve (conductance method.)

III.1.1 Conductance Method

The conductance method, developed by Nicollian and Goetzberger (42), yields highly accurate and reliable results. As low as few times $10^9$ states/cm$^2$-eV can be detected. The information about interface states is extracted from the conductance of the MIS capacitor measured with an ac-signal of 5 - 100 kHz. There exists a measurable conductance because interface states cannot follow the measuring signal immediately. They lag behind. In one-half of the cycle the electrons in the silicon are at a higher energy than the electrons in the interface states. Therefore, electrons are captured by interface states at a lower average energy and they lose energy. In the other half of the cycle the electrons are at a higher average energy than the electrons in the silicon. As the electrons are emitted from the interface states into the silicon they lose energy again. Thus, there will be an energy loss on both halves of the cycle which must be supplied by the signal source. The measured conductance depends on the density and energy distribution of the interface states, the measuring frequency and the bias voltage of the MIS diode. By measuring conductance as a function of bias with a fixed frequency and as a function of frequency with a fixed bias voltage we can obtain enough information to be able to extract the density of interface states and their time constants (capture cross-sections).

This method is most accurate in the weak inversion range. For the depletion range the method can still be used but a much more
complicated model has to be assumed. Random fluctuations of the surface potential due to randomly distributed impurities and the time constant dispersion become important. In the strong inversion and accumulation regions, however, the method is not valid any more because either the majority carriers or the minority carriers (for accumulation or inversion, respectively) are in good contact with the corresponding band (conductance or valence band). The time constants are, therefore, very short and no conductance can be measured even for the highest reasonable measurement frequencies.

The limitations of the conductance method are, therefore, that it is quite complicated to instrument, that the analysis is time consuming and that with one slice only a limited fraction of the band gap can be investigated. In order to cover the upper and lower half of the band gap n- and p-type samples have to be used. Near the band edges the method cannot be used at all. The advantages of the method are the great accuracy and sensitivity.

III.1.2. Capacitance Method

The evaluation of interface state density using the capacitance method can be achieved by three different procedures:

a) The differentiation procedure
b) The integration procedure
c) The temperature procedure.

a) Differentiation procedure

This method has been used first by Terman(47). The MIS capacitance as a function of applied bias voltage is measured at a high
frequency ($\approx 1$ MHz) such that the interface states cannot follow. The interface states together with the surface charge cause a shift of the experimental C-V curve with respect to the theoretical* C-V curve. This shift, $\Delta V$, is a function of surface potential. If we define surface energy $E$ as the energy difference in eV between the Fermi level and the upper edge of the valence band (see Fig. 3.1), $E$ is related to the surface potential $u_s$ by:

$$E = (kT/q)(u_s - u_F) + E_g/2$$  \hspace{1cm} (3.1)

where $u_s$ and $u_F$ are the surface potential and the Fermi potential respectively measured in kT/q units with respect to the intrinsic Fermi level in the bulk (see Fig. 3.1 and the list of symbols). $E_g$ is the band gap energy in eV. The voltage shift $\Delta V$ of the experimental and the theoretical C-V curve is related to the surface charge $Q_{ss}$ and the interface state density $N_{ss}$ by

$$Q_{ss} + \int_{0}^{E} N_{ss}(E') dE' = C_i \Delta V$$  \hspace{1cm} (3.2)

where $C_i$ is the insulator capacitance per unit area. The effect of any work function differences are neglected. Differentiation of (3.2) with respect to $E$ gives:

* The theory of the MIS system will not be given here. For a review see for example Grove (24), Chapter 9.
\[ qN_{ss}(E) = C_i \frac{\Delta V}{\partial E} = \left( \frac{qG}{kT} \right) \left( \frac{\partial \Delta V}{\partial u_s} \right)_V. \] (3.3)

The last step in (3.3) follows from (3.1). \( \Delta V \), which is measured as a function of \( V \) has to be converted into a function of \( u_s \) by using the well known relationship between \( V \) and \( u_s \) for an ideal MIS capacitor as given, for example, by Goetzberger\(^{48} \). The interface state density per unit energy is, then, given by:

\[ N_{ss}(E) = \left( \frac{C_i}{kT} \right) \left( \frac{\partial \Delta V}{\partial u_s} \right)_V. \] (3.4)

\( \left( \frac{\partial \Delta V}{\partial u_s} \right)_V \) is obtained by graphical differentiation of the measured voltage shift \( \Delta V \) with respect to \( u_s \). This step imposes severe limitations on the accuracy of the method as pointed out by Zaininger and Warfield\(^{49} \). Recently Sah et al.\(^{50} \) have shown that, over a limited part of the energy gap (\( \pm 14 \) kT at \( 300^\circ \)K centered at midgap), the method is sensitive to \( 2 \times 10^9 \) states/cm\(^2\)-eV if the measurement accuracy is pushed to its limits and the full capacity of the computer is used in the analysis.

b) Integration Procedure

Berglund\(^{51} \) proposed a method of obtaining interface state density over an extended energy range using a low frequency thermal equilibrium C-V curve. This method has the advantage that no graphical differentiation has to be performed. Berglund showed that

\[ \frac{du_s}{dV} = \left( \frac{q}{kT} \right) \left[ 1 - \frac{C(V)}{C_i} \right] \] (3.5)
where \( C(V) \) is the MIS capacitance as a function of bias voltage \( V \).

Integrating (3.5) we obtain experimentally the surface potential \( u_s \) as a function of applied voltage \( V \) up to an additive constant:

\[
u_s = \left(\frac{q}{kT}\right) \int_{V_{acc}}^{V} \left[1 - \frac{C(V)}{C_i}\right] dV + \Delta
\]

where \( V_{acc} \) is any voltage corresponding to strong accumulation and the integration is performed toward inversion. Combining (3.5) and (3.6) \( du_s/dV \) can be plotted as a function of the surface potential difference \((u_s - \Delta)\). Berglund found the unknown constant \( \Delta \) by matching this experimental \( du_s/dV \) vs. \((u_s - \Delta)\) curve to a theoretical \( du_s/dV \) curve for values of \( u_s \) corresponding to strong accumulation. In strong accumulation the charge in interface states is much smaller than the total space charge in the semiconductor so that \( \Delta \) can be found within a small fraction of \( kT/q \). Once \( u_s \) is known as a function of \( V \), the interface state density is given by (51):

\[
N_{ss}(u_s) = \left(\frac{1}{q}\right) \left\{ \frac{C[u_s(V)]}{1 - C[u_s(V)]/C_i} - C_{sc}(u_s) \right\}
\]

\( C_{sc}(u_s) \) is the theoretical incremental low frequency semiconductor surface capacitance including the inversion layer response.

It should be noted that it is imperative in Berglund’s method that the C-V curve is a true low frequency thermal equilibrium curve.
Otherwise the equations given are not valid. This is quite difficult to achieve. Berglund has observed low frequency C-V dispersion down to 1.5Hz using a lock-in technique. Below this frequency conventional sinusoidal phase sensitive measurement techniques become difficult to perform and, therefore, this method has not been widely used.

Solutions to the problem of obtaining a thermal equilibrium C-V curve have been given by Castagne\(^{(52)}\)-\(^{(54)}\) and Kuhn\(^{(55)}\). They have shown that the incremental thermal equilibrium MIS-capacitance \(C(V)\) is proportional to the charging current if a linear voltage ramp is applied to an MIS capacitor, i.e.,

\[
C(V) = \frac{i(V)}{\alpha} \tag{3.8}
\]

where \(i(V)\) is the charging current and \(\alpha\) the slope of the applied voltage ramp. With this technique true thermal equilibrium C-V curves can be obtained because the slope of the voltage ramp can be made much smaller than the equivalent rate of change of the ac-signal at the lowest possible frequency in Berglund's technique. In contrast to Berglund's approach, Kuhn determined the constant \(\Delta\) by matching an experimental \(C-(u_s - \Delta)\) curve with a theoretical \(C-u_s\) curve in strong accumulation and strong inversion. This should give better accuracy than Berglund's approach. Castagne calculated \(\Delta\) theoretically.

In this thesis interface states are measured using Kuhn's method. The most time consuming step in Kuhn's technique is the analysis of the experimental data to obtain the desired interface state densities.
To facilitate the analysis a computer program (FORTRAN IV) has been written which carries out the complete analysis including the determination of \( \Delta \) automatically. This program is described in a separate publication\(^{56}\). The low frequency C-V curves are obtained on an experimental apparatus identical to the one described by Kuhn. Details of the experimental setup and the theory of measuring low frequency capacitance-voltage curves can be found in Ref. (55).

Two examples of measured C-V curves obtained from MIS capacitors fabricated on nominal 3-7 ohm-cm n-type silicon substrates are shown in Figure 3.2. These capacitors had the double insulator structure (500 Å Al\(_2\)O\(_3\), 1000 Å SiO\(_2\)) described in Chapter II. The solid curve in Figure 3.2 was obtained from a capacitor which had a 20 mil diameter evaporated Al dot as the field plate while the broken curve was obtained from a capacitor which had the standard 10 mil square Ti-Al dot as the field plate. These C-V curves were obtained at a sweep rate of 0.2 volts/sec and a current level of approximately 100 and 20 picoamperes, respectively. To improve the accuracy the recorded current versus voltage curve was only used to find \( C / C_i \) and the insulator capacitance \( C_i \) was measured in strong accumulation using a standard capacitance bridge. The only additional data required for an analysis are the field plate area and the doping density at the semiconductor-insulator interface. The field plate area was measured with a microscope and the doping density was determined from a high frequency C-V curve as will be described in section 3.4. These low frequency C-V curves were then used to determine interface state density on a number of different slices.
Fig. 3.2 MIS capacitance-voltage characteristics obtained with the quasi-static technique for a double insulator structure as shown in the insert.

a) Slice #TC8, \( N_D = 7.75 \times 10^{14} \) cm\(^{-3} \),
field plate area = \( 2.02 \times 10^{-3} \) cm\(^2 \) (20 mil dia. dot),
insulator: 1000Å SiO\(_2\) + 500Å Al\(_2\)O\(_3\)

b) Slice #03111, \( N_D = 1.77 \times 10^{15} \) cm\(^{-3} \)
field plate area = \( 6.45 \times 10^{-4} \) cm\(^2 \) (10 mil square)
insulator: 1000Å SiO\(_2\) + 500Å Al\(_2\)O\(_3\)
As an example the interface state density as a function of surface energy (see Fig. 3.1) obtained from a device with the standard double insulator structure (500 Å Al₂O₃, 1000 Å SiO₂) is compared in Fig. 3.3 to that obtained from a device with a single insulator (1000 Å SiO₂). Both curves show a minimum at essentially midgap with the curve for SiO₂ slightly higher and both curves show an increase towards the band edges. This picture is consistent with the results obtained for the Si-SiO₂ interface using the ac conductance technique (42).

In Fig. 3.4 a comparison is made between a steam grown and a dry SiO₂ oxide. Around midgap the interface state density of the dry oxide is approximately 10¹⁰ states/cm²-eV smaller than that obtained from the steam oxide. The wet oxide seems to have a high peak in interface state density near the valence band. However, the inaccuracy of the measurement method increases from midgap to the band edges so that this finding has to be interpreted with caution.

c) Temperature Procedure

Interface state density near the majority band edge may be determined with the Gray - Brown (57) technique in which the change in flatband voltage of an MIS capacitor is measured as a function of temperature. The range of band gap accessible is very small, being determined by the variation of the Fermi level within the temperature limits dictated by oxide instabilities at high temperatures, and impurity ionization at low temperatures. In addition, this technique is not easy to instrument and the accuracy is poor, only of the order of 10¹² states per cm²-eV.
Fig. 3.3 Interface state density as a function of energy in the forbidden band gap (with the upper edge of the valence band taken as reference) for

a) MIS capacitor with a double insulator structure (1000Å SiO₂ + 500Å Al₂O₃)

b) MIS capacitor with a single insulator structure (1000Å SiO₂)
Fig. 3.4 Interface state density as a function of energy in the forbidden band gap for

a) MIS capacitor with a double insulator structure 
   (1000Å SiO₂ + 500Å Al₂O₃) where the SiO₂ is a dry oxide.

b) MIS capacitor with the same insulator structure as in
   a) but a wet SiO₂ layer.
Summarizing the discussion about interface state density measurements, the ac conductance method\(^{(42)}\) yields probably the most detailed and accurate information. But it is very laborious and only a small fraction of the band gap can be probed conveniently. (between midgap and Fermi level). The energy region near the band edges is not accessible. The high frequency capacitance method is too inaccurate for interface state density measurements (at least with a reasonable effort). The low frequency, thermal equilibrium C-V method\(^{(51)-(56)}\) with Kuhn's\(^{(55)}\) experimental technique and computer aid for fast analysis\(^{(56)}\) is probably the optimal method reported so far. The only place the temperature method of Gray and Brown\(^{(57)}\) could be used is to obtain an estimate of the interface state density very near the band edges where all the above methods are inaccurate.

### III.2 Flatband Voltage

If no voltage is applied to the gate of an MIS structure, the energy bands at the semiconductor surface are usually bent upwards or downwards (upwards for the double insulator structures used in this work) as shown in Fig. 3.5a, resulting in a space charge region at the semiconductor surface. The physical effects which give rise to this initial band bending are:

a) Work function difference between metal and semiconductor,

b) Fixed charge in the insulator,

c) Fixed surface charge,

d) Interface states.
Fig. 3.5 Band diagram of a real MIS structure

a) Condition for $V_G = 0$

b) Flatband condition
The work function difference $\Delta W_{MS}$ is a well known quantity for the Metal-SiO$_2$-Si system. It depends on the doping density of the semiconductor, being about -0.7V and zero for p- and n-type silicon doped with $10^{16}$ cm$^{-3}$ impurities, respectively (58). For the double insulator structures, however, no reliable value is known. The additional contact potential difference which seems to exist at the Al$_2$O$_3$-SiO$_2$ interface (19), (59) further obscures the picture. For the following investigations it is, however, not necessary to know $\Delta W_{MS}$, because it is measured together with the fixed oxide charge and the surface charge and there is no need to separate them.

The influence of charges within the insulator on the initial band bending can easily be understood. Any charge in the insulator induces a mirror charge in the semiconductor depending on its magnitude and the distance from the interface. As mentioned in Section III.1, the fixed charges in the insulator are lumped together with the fixed surface charge by adjusting their magnitude in such a way that the band bending stays the same.

The flatband voltage $V_{FB}$ is now defined (24) as the gate voltage required to bring about a flatband condition as illustrated in Fig. 3.5b. In other words $V_{FB}$ is the gate voltage required to remove the space charge at the semiconductor surface. A derivation of $V_{FB}$ as a function of the IGFET parameters has been given by various authors.* In this thesis this relationship is slightly modified to be consistent with the classification of surface states introduced in Section III.1:

* See, for example, Grove (24), p. 281.
\[ V_{FB} = \phi_H^i - Q_{ss}/C_i + (q/C_i) \int_0^E N_{ss}(E') dE' \]  
(3.9)

where \( E \) is the separation of the Fermi level from the valence band, given by (3.1) with \( u_s = 0 \), i.e.,

\[ E = E_g/2 - (kT/q)u_F \quad (u_F < 0) \]  
(3.10)

The last term in (3.9), the influence of the interface states on the flatband voltage, is usually neglected in the literature (24), (60).

Flatband voltages can best be measured by a comparison of a high frequency (≈1 MHz) C-V curve and an ideal C-V curve (48). The voltage shift of the experimental relative to the ideal curve at the flatband capacitance level corresponds to the flat-band voltage as defined in this chapter. The flat band capacitance level is the value of the ideal capacitance for zero surface potential (48). This procedure is illustrated in Fig. 3.6.

Using the flatband voltage \( V_{FB} \) as given in (3.9) has the disadvantage that part but not all of the interface states are included in \( V_{FB} \). Those up to the Fermi level are included, those above are not. If special emphasis is put on the influence of interface states on the device characteristics, as will be done in Chapter V, where the transition region of an IGFET is investigated, it is better to exclude interface states from (3.9) completely and use

\[ V_0 = \phi_{MS} - Q_{ss}/C_i \]  
(3.11)
Fig. 3.6 Method of finding the flatband voltage $V_{FB}$ from a high frequency C-V curve by comparing it to a theoretical high frequency C-V curve.
to characterize the device. Of course, \( V_o \) is not the flatband voltage anymore because it does not agree with the definition of flatband voltage. Only if interface states can be neglected compared to \((\frac{1}{M^*} - Q_{SS}/C_i)\) is \( V_o \approx V_{FB} \). \( V_o \) is a convenient parameter to characterize an IGFET because it is a constant and does not depend on temperature, doping density, etc. (at least to first order). The effect of interface states on the device performance can be included separately by measuring them with one of the methods outlined in Section III.1.

A new method of measuring \( V_o \) which is used in this thesis will now be presented.

In contrast to the standard method of measuring \( V_{FB} \) (see Fig. 3.6), \( V_o \) is deduced from the small signal channel conductance \( g_D \) versus gate voltage \( V_G \) characteristic of an IGFET.* A typical example of such a characteristic is shown in Fig. 3.7. \( g_D(V_G) \) is measured point by point with a Boonton conductance bridge as shown in the insert to Fig. 3.7. The measuring ac-voltage applied between source and drain had a frequency of 5kHz and a peak-to-peak value of approximately 1mV. The influence of the ac-drain voltage on the channel conductance will be treated in detail in Chapter V and the Appendix. No dc-bias was applied between drain and substrate and the source was tied to the substrate. \( g_D(V_G) \) has a typical S-shape, the lower bend belonging to the transition region which will be treated in Chapter V and the upper bend corresponding to the region where surface scattering of the carriers in the channel increases strongly. In an intermediate range

---

* The gate voltage \( V_G \) is always understood to be the voltage between gate and source. (See also List of Symbols).
Fig. 3.7 Dependence of the small signal channel conductance on the gate voltage $V_G$. The insert shows the measurement arrangement.
$g_D(V_G)$ is fairly linear and can be represented by:

$$g_D = - (W/L) \mu_p C_i (V_G - V_T) \quad (3.12)$$

for a p-channel device. $W/L$ is the width to length ratio of the channel (see Fig. 2.1), $\mu_p$ the hole mobility and $V_T$ the intercept of the extrapolated linear region of $g_D$ with the voltage axis. $\mu_p$ is nearly constant in the linear range of $g_D(V_G)$ as will be shown in detail in Chapter V. $V_G$ and $V_T$ are both negative for enhancement mode p-channel devices. Equation (3-12) is a first order equation but in the range where it is used (linear part of $g_D(V_G)$) it is quite accurate. The conductance values are large enough so that interface state trapping, etc., have negligible influence and are low enough so that mobility reduction need not be included.

To find $V_o$ we have to calculate first the mobile hole charge density $Q_p$ in the channel of an ideal IGFET (no surface states or work function difference).

$Q_p$ is a function of surface potential $u_s$ and is given by:

$$Q_p(u_s) = q n_i L_D \int_{u_F}^{u_s} \frac{\exp(u_F - u) \, du}{\left[2\cosh(u_F - u) - \cosh(u_F) + u \sinh(u_F)\right]^{1/2}} \quad (3.13)$$

where $n_i$ is the intrinsic carrier concentration and $u$, $u_F$, and $u_s$ are the electrostatic potential, the Fermi potential and the surface potential respectively, measured in kT/q units with respect to the

* See for example Grove(24), p. 324.
intrinsic Fermi level in the bulk. \( L_D \) is the intrinsic Debye length given by

\[
L_D = \sqrt{\frac{kT}{q}} \left( \frac{\varepsilon_s}{2q\epsilon_i} \right)
\]  

(3.14)

where \( \varepsilon_s \) is the dielectric constant of the semi-conductor.

For gate voltages well above threshold the gate voltage \( V_G \) and the surface potential \( u_s \) for an ideal IGFET are related by:

\[
V_G = \left( \frac{kT}{q} \right) u_s - \frac{Q_s}{C_i}
\]  

(3.15)

where \( Q_s(u_s) \) is the total induced charge per unit area in the semi-conductor given by:

\[
Q_s(u_s) = 2q\epsilon_i L_D \left\{ 2 \left[ \cosh(u_F - u_s) - \cosh(u_F) + u_s \sinh(u_F) \right] \right\}^{1/2}
\]  

(3.16)

Combining equations (3.13)-(3.16) \( Q_p \) can be calculated as a function of \( V_G \). A typical \( Q_p(V_G) \) curve is shown in Fig. 3.8. Above a certain value of \( Q_p \) denoted by \( Q_p^* \), \( Q_p \) vs. \( V_G \) is a straight line. This is obvious because for strongly inverted surfaces virtually all the induced charge is mobile charge and the depletion charge is constant. Below \( Q_p^* \), for weakly inverted surfaces, the depletion region is still widening and part of the induced charge is immobile space charge.

The portion of \( Q_p(V_G) \) above \( Q_p^* \) can be represented by:
Fig. 3.8 Mobile charge density as a function of gate voltage for an ideal structure.
\[ Q_p = C_i (V_G - V'_T) \]  \hspace{1cm} (3.17)

where \( V'_T \) is the intercept of the extrapolated linear region of \( Q_p \) with the voltage axis. We can find \( V'_T \) by calculating \( Q_p(u_s) \) from (3.13) and (3.14) using the appropriate values for \( u_F \) and \( C_i \). With the help of (3.15) and (3.16), \( Q_p(u_s) \) is converted into \( Q_p(V_G) \). Extrapolating the \( Q_p(V_G) \) curve above a reasonably chosen \( Q_p^* \) to the intercept with the \( V_G \)-axis yields \( V'_T \).

\( V_0 \) from equation (3.11) is now given by

\[ V_0 = q'_{HS} - Q_{ss}/C_i = V_T - V'_T \]  \hspace{1cm} (3.18)

Equation (3.18) needs some explanation. \( V_T \) in eq. (3.12) was obtained from conductance values well above threshold where the silicon surface is strongly inverted. The band picture for the lowest conductance value used for the straight line approximation is shown in Fig. 3.9. The Fermi level corresponding to this conductance value is close to the valence band edge and, therefore, only a very small part of the interface states are occupied as indicated in Fig. 3.9. The deviation of \( V_T \) from \( V'_T \) —the voltage corresponding to \( V_T \) for an ideal structure—is, therefore, equal to the flatband voltage minus the contribution of the interface states. \( V'_T \) can be found, for example, from equation (3.17). Thus, equation (3.18) is proved.

To perform the calculation of \( V_0 \) automatically a computer
Fig. 3.9 Band diagram for an IGFET in strong inversion ($g_0 = g_2$ in Fig. 3.7). The shaded area represents the portion of the interface states which are occupied.

Fig. 3.10 Comparison of flatband voltage $V_{FB}$ and $V_0$ measured on a number of double insulator MIS capacitors and IGFETs, respectively.
program (FORTRAN IV) has been written.* The program starts with the
device parameters $N_D$ (doping density), $C_1$, $W$, $L$ and a measured $g_D(V_G)$
curve in digital form. All subsequent steps are performed automati-
cally without the need of intermediate graphing. A critical step in
the analysis is deciding which points of the $g_D(V_G)$ curve to use for
the straight line approximation. In the program this decision is
made by taking only those $g_D$ values located between an upper and
lower bound, labeled $g_1$ and $g_2$ in Fig. 3.7. $g_1$ and $g_2$ are chosen by
graphing one $g_D(V_G)$ curve and determining the conductance levels at
which the curve starts to deviate from a straight line. These limits
are then valid for all IGFETs with the same (W/L) ratio and approxi-
mately the same oxide capacitance as can be seen from equation (3.12).
For the FET 3 (W/L = 10) used in this work the best choice for $g_1$
and $g_2$ turned out to be 30 $\mu$mhos and 9 $\mu$mhos respectively. For the
FET 4 (W/L = 20) $g_1$ and $g_2$ were taken as 60 $\mu$mhos and 15 $\mu$mhos re-
spectively.

Values of $V_o$ obtained on a number of IGFETs are now compared
to the flatband voltages, $V_{FB}$, determined from high frequency C-V
curves. The result is given in Fig. 3.10. The MIS capacitors chosen
to measure $V_{FB}$ are in the immediate vicinity of the IGFET where $V_o$ is
measured to ensure identical surface properties. It can be seen from
Fig. 3.10 that for some slices $V_o$ and $V_{FB}$ are practically identical
but that generally $V_o$ is larger than $V_{FB}$ indicating the influence of
interface states. It should be noted, that all slices have extremely
low interface state densities, of the order of $10^9$ states/cm$^2$-eV

* A listout of the program is available on request.
at midgap. In spite of this, differences in $V_o$ and $V_{FB}$ can be seen. The assumption that interface states can be neglected in the expression for $V_{FB}$ (equation (3.9)) is, therefore, not always justified.

III.3. Threshold Voltage

Perhaps the most important parameter of an IGFET for the applications engineer is the threshold voltage $V_T$. It is defined as the gate voltage which has to be applied to an IGFET to form a conducting channel. For p-type enhancement mode devices $V_T$ is negative. Despite its practical importance it will only be covered briefly in this work. Detailed discussions are given by Grove, Brotherston, and Richman among others.

The threshold voltage cannot be defined very precisely. If the gate voltage is increased the IGFET undergoes a gradual transition from a nonconducting state (depletion mode) to a conducting state (inversion mode). Defining the onset of conduction must, therefore, necessarily be ambiguous.

The following definitions of threshold voltage exist in the literature and are in practical use:

a) $V_T$ is the gate voltage where the drain current has reached a certain value, e.g., 0.1 $\mu$A, 1 $\mu$A, 10 $\mu$A, etc.

b) $V_T$ is the intercept of the square root of the saturation drain current versus gate voltage characteristic with the gate voltage axis.

c) $V_T$ is the gate voltage where the small signal conductance has reached a certain value, e.g., 0.01 $\mu$mhos, 0.1 $\mu$mhos, etc.
d) $V_T$ is the intercept of the linear part of the small signal conductance with the gate voltage axis. This method corresponds to the procedure described in Section III.2 so that $V_T$ in equation (3.12) can be interpreted as the threshold voltage.

The method of measuring threshold voltage is usually chosen from practical consideration. All the definitions mentioned above give slightly different values. A severe limitation is that an exact relationship between the threshold voltage and the corresponding band bending (surface potential) cannot be given. Grove$^{(24)}$ assumes that the surface potential at threshold is twice the Fermi potential, i.e.,

$$
u_s(\text{threshold}) = 2\nu_F$$  \hspace{1cm} (3.19)

This is only approximately true and depending on the method chosen to measure $V_T$ can deviate considerably.

We have checked the validity of (3.19) for the case that the threshold voltage $V_T$ is defined as the intercept of the linear part of the conductance curve with the gate voltage axis. From (3.15) we have calculated the surface potential $\nu_s$ belonging to a gate voltage $V_G = V_T$. This was done for a number of devices and the results are presented in Fig. 3.11. The curve in Fig. 3.11 indicates that the factor 2 in (3.19) is better replaced by 2.18 for this threshold voltage definition; i.e., equation (3.19) is approximately 10 percent off.

Summarizing the above remarks, $V_T$ cannot be defined without specifying the method by which it is measured. Once a method is
Fig. 3.11 Ratio of surface potential $u_s$ and Fermi potential $u_F$ at threshold. ($V_T$ defined as intercept of linear part of $g_D(V_G)$ with $V_G$-axis; see Fig.3.7)

Fig. 3.12 Experimental method used for transient measurements. (After A. Waxmann(14)).
chosen threshold voltage changes can best be used to describe the device behavior.

In our work the use of $V_T$ as a device parameter has been avoided. The limitations and ambiguities mentioned above are especially important in or near the transition region of the IGFET which will be covered in Chapter V. The flatband voltage $V_{FB}$ and especially $V_o$ are better suited to characterize a device if the gradual transition region at the onset of conduction is to be investigated. $V_{FB}$ and $V_o$ are defined in terms of well conceived physical quantities such as $Q_{SS}$, $C_i$, $I_{MS}$, etc. (See Section III.2), and no ambiguous approximations or assumptions have to be made. This does not, however, violate the initial statement that the threshold voltage is probably the most important IGFET parameter for the applications engineer. In circuit applications a crude but simple model is usually quite satisfactory and uncertainties due to an ambiguous definition of $V_T$ are unimportant.

III.4 Surface Doping Density

An IGFET parameter which enters practically all device calculations is the doping density at the semiconductor surface. It determines essentially the Fermi potential $u_F$ which is given for an n-type semiconductor by:

$$u_F = -\ln\left(\frac{N_D}{n_i}\right)$$  \hspace{1cm} (3.20)
where \( N_D \) is the doping density in \( \text{cm}^{-3} \). Differences in the effective masses of holes and electrons are neglected in (3.20).

It is well known that the impurities near the surface of a silicon substrate redistribute during oxidation \((62),(63)\). It is, therefore, usually not admissible to use the starting substrate resistivity to calculate the surface doping density. As shown by Deal et al. \((63)\) phosphorus which was used as dopant in the experimental n-type structures described here can pile up at the surface by more than 50% during oxidation depending on oxidation temperature and ambient. For the oxidation conditions which were chosen for the experimental structures used in this work the pile up should be about 10%. To obtain a reliable value, the surface doping density should, therefore, be measured on the oxidized slice.

Several methods of determining the surface doping density have been described in the literature and are in practical use. These methods with their advantages and limitations will now be described.

### III.4.1 Surface Doping Density Determined From High Frequency C-V Curve.

The surface doping density can be calculated from the ratio of maximum to minimum capacitance of a high frequency C-V curve, measured on a MIS capacitor (see Fig. 3.6). Lindner \((63)\) has shown that:

\[
\frac{C_{\text{max}}}{C_{\text{min}}} = 1 + \left( \frac{C_i}{q} \right) \sqrt{\frac{2kT}{\varepsilon_s N_D}} \frac{U_i}{\sqrt{U_i - V}}
\]

\[(3.21)\]
where $C_{\text{max}}$ and $C_{\text{min}}$ are the maximum and minimum measured capacitance as shown in Fig. 3.6. $\varepsilon_s$ is the dielectric constant, for silicon $= 1.06 \times 10^{-12}$ Farads/cm. $u_1$ is a particular potential defined by the following equation:

$$u_1 - l = \exp(u_1)(N_D/n_i)^2$$  (3.22)

For a given $C_{\text{max}}/C_{\text{min}}$ ratio, $N_D$ can be calculated by solving (3.21) together with (3.22) numerically. $N_D$ decreases approximately as the one-half power of $C_{\text{max}}/C_{\text{min}}$.

Various approximations to the above approach have been given. Deal et al. (63) approximate the last factor in equation (3.21) by $\sqrt{\frac{4}{\ln(N_D/n_i)}}$ so that (3.21) is replaced by:

$$C_{\text{max}}/C_{\text{min}} = 1 + \left(\frac{C_i}{q}\right)\sqrt{2kT/(\varepsilon_sN_D)} \sqrt{\frac{4}{\ln(N_D/n_i)}}$$  (3.23)

This approximation is based on the assumption that the width $x_d$ of the depletion region reaches its maximum value $x_{d\text{max}}$ when the hole surface concentration $p_s$ in the inversion region equals the doping concentration in the bulk of the semiconductor, i.e., $p_s = N_D$ and stays constant for stronger inversion (65).

A different approximation was used by van Gelder (66). Writing $x = \ln C_{\text{max}}/(C_{\text{max}}/C_{\text{min}} - 1)$, a function $N_D = N_D(x)$ can be generated using (3.21) and (3.22). Van Gelder approximated this function by a fifth order polynomial. Once the polynomial coeffi-
cients are determined \( N_D \) can be calculated easily for any value of \( x \). Van Gelder's values of \( N_D \) are within a few percent of the more exact but more complicated calculation based on equation (3.21) for all practical cases.

All calculations of \( N_D \) from the \( C_{\text{max}}/C_{\text{min}} \) ratio of the high frequency curve are best carried out with the help of a computer. As an alternative Pierret and Sah\(^{67}\) developed a slide rule for computing the surface doping density from the measured \( C_{\text{max}}/C_{\text{min}} \) ratio.

The high frequency C - V method is certainly the most widely used method for determining surface doping density. It is easy to instrument and if a computer is readily available, fast to analyse. The accuracy of the method is approximately 20%. This rather poor accuracy is due to the following limitations:

1) Very often the MIS capacitance in inversion is not constant. It can increase slightly for stronger inversion, particularly in p-type substrates\(^{68}\), or it can even show hysteresis due to trapping instabilities, ion separation on the oxide surface, etc. It is, therefore, often difficult to assign a correct value to \( C_{\text{min}} \).

2) The theoretical expressions (3.21) - (3.23) are based on a uniform impurity distribution in the silicon. As mentioned before and shown by Deal et al.\(^{63}\) the doping density \( N_D \), however, is usually a function of distance normal to the semiconductor surface so that the value of \( N_D \) calculated from the \( C_{\text{max}}/C_{\text{min}} \) ratio represents some average concentration across the depletion layer. Therefore, the true value of \( N_D \) in the immediate vicinity of
the surface will be somewhat higher than the calculated value (for phosphorus doped silicon).

Very often the high frequency C-V method is accurate enough. For other applications, however, better accuracy is needed. In the next section some alternative methods will be described. Some of them can give a better accuracy, but are, usually, more complicated to instrument, to analyze or in sample preparation.

In this thesis van Gelder's method (66) was usually employed. The high frequency C-V curves were measured with a standard experimental setup composed of a Boonton Electronics Corp. Model 71A capacitance meter, an electronic sweep circuit, an appropriate test fixture and an xy-recorder. Fig. 3.6 shows a block diagram of the measuring system and a high frequency C-V curve of an MIS capacitor plotted by this equipment. A simple TELCOMP program* was then used to calculate the surface doping \( N_D \) from the measured \( C_{\text{max}}/C_{\text{min}} \) ratio. The same high frequency C-V curves were used to determine the flat-band voltage as described in Section III.2.

All high frequency C-V curves were taken on the 10 x 10 mils MIS capacitors located in the center row of a slice (see Chapter II). The error made by taking the results obtained on these capacitors to infer the doping density under the gate of an adjacent IGFET is negligible since the doping density is highly uniform parallel to the surface of the slice (63). To avoid drifting of the C-V curves due to the charge separation effect (69), (70), a large positive voltage, e.g. +50 volts, was applied to the capacitor immediately before

* A listing of this program is available from the author.
taking the measurement. This positive bias attracts any negative mo-
bile ions on the outer surface of the oxide to the metal plate and re-
duces their effect on the MIS inversion layer for the time of the
measurement. During the measurement the slices were kept in complete
darkness.

III. 4.2 Alternative Methods for Determining Surface Doping Density

Grove et al.\(^{(58)}\) have shown that if the capacitance of an MIS
capacitor is measured under such condition that the minority carriers
cannot accumulate near the surface even in the bias range corresponding
to inversion, i.e., under transient condition resulting in a deep de-
pletion C-V curve, the MIS capacitance is given by:

\[
\frac{C}{C_i} = \left(1 + \frac{V_G'}{\alpha^2}\right)^{-\nu_2} \tag{3.24}
\]

where \(V_G' = V_G - V_{FB}\) is the effective gate voltage, \(V_{FB}\) is the flat-
band voltage (see Section III.2) and

\[
\alpha^2 = \frac{1}{C_i^2} \left(q N_D \varepsilon_S / 2\right) \tag{3.25}
\]

According to (3.24) a plot of \((C_i/C)^2\) versus \(V_G\) should yield
a straight line if \(N_D\) is constant. \(N_D\) can be determined from the
slope \(\alpha^2\) of this straight line by solving (3.25). If the doping
density is a function of distance normal to the interface the exper-
imental \((C_i/C)^2\) versus \(V_G\) plot will deviate from a straight line.
The slope of this curve gives an indication of the impurity profile in the semiconductor. The accuracy, however, is usually not too good, since differentiation of an experimental curve is highly unreliable.

A convenient way of obtaining depletion type C-V curves has been described by Kuhn(71). A sawtooth voltage of approximately 10 kHz is applied to an MIS capacitor driving it periodically into deep depletion without accumulation of minority carriers at the surface. The charging current is proportional to the MIS capacitance and can be displayed on an oscilloscope. Kuhn used an intermediate sampling unit and recorded the deep depletion C-V curve directly on an xy-recorder.

Doping densities determined from deep depletion C-V curves are accurate within approximately \( \pm 10\% \) for devices where the pile up or depletion of impurities at the surface is small, i.e., where the doping density is nearly uniform. The accuracy is better or at least as good as for the high frequency C-V method described before.

Another way of determining the surface doping density is by measuring the capacitance \( C \) of a Schottky diode as a function of reverse bias voltage \( V \) and use of the equation:

\[
N_D(x) = \left( q\varepsilon S A^2 \right)^{-1} C^3 / (dC/dV)
\]  

(3.26)

where \( A \) is the area of the diode(72). The depth \( x \) is found from the measured capacity, \( x = \varepsilon_S A/C \). This method is mostly used to find the impurity profile in epitaxial layers. To use it for determining the
surface doping density of an oxidized silicon wafer, the oxide layer has to be removed from part of the slice and then Schottky diodes have to be formed. Since that does not involve high temperature steps (>500°C) the doping profile should stay the same.

The advantage of this method is the significant distance over which the doping density can be obtained—essentially from the surface to a depth of several microns, limited only by the onset of avalanche breakdown in the Schottky diode. The disadvantage of the Schottky diode method is the limited numerical accuracy which comes from the necessity of interpolating between two points to find dC/dV in (3.26). Another limitation, especially for routine testing of oxidized slices, is the complicated preparation of the samples to be measured.

An improved version of the Schottky diode method has been reported recently by Copeland\(^73\). This technique involves driving a Schottky diode with a small constant RF current (a few hundred \(\mu\)A at 5 MHz). The depth \(x\) of the depletion layer is varied by changing the dc bias, and the inverse doping profile \(N_D^{-1}(x)\) is obtained by monitoring the voltage across the diode at the fundamental frequency, which is proportional to the depth \(x\), and the second harmonic voltage, which is proportional to \(N_D^{-1}\). This method has an improved accuracy (error is < 10% depending on \(x\)) and no intermediate calculations are necessary. \(N_D^{-1}(x)\) can be plotted directly on a recorder.

Waxman\(^74\) described a method of determining surface doping density using the transient behavior of an MIS capacitor. A narrow voltage pulse \(\Delta V_G\) is applied to an MIS capacitor in the configuration shown in Fig. 3.12. The change in the capacitor charge, \(\Delta Q\), is
measured with an oscilloscope. For any surface potential $u_s$ the depletion charge of the MIS capacitor is given by:

$$Q_D = \sqrt{2\varepsilon_S N_D kT \sqrt{|u_s|}}$$  \hspace{1cm} (3.27)

If the voltage pulse is narrow enough only the majority carriers are able to follow the signal resulting in a widening or narrowing of the depletion region. The depletion charge after application of the pulse is:

$$Q_D' = \sqrt{2\varepsilon_S N_D kT \sqrt{|u_s'|}}$$  \hspace{1cm} (3.28)

where $u_s'$ is the surface potential corresponding to the new depletion charge. The measured change in the MIS capacitor charge $\Delta Q$, is given by:

$$\Delta Q = Q_D' - Q_D = \sqrt{2\varepsilon_S kT N_D} \left(lu_s'^{1/2} - |u_s|^{1/2} \right)$$  \hspace{1cm} (3.29)

where (3.27) and (3.28) have been used, because the depletion charge is the only charge component which can change under transient conditions. An alternative expression for $\Delta Q$ can be obtained from:

$$\Delta Q = C_i \Delta V_i$$  \hspace{1cm} (3.30)
where $\Delta V_i$ is the change in voltage drop over the oxide. $\Delta V_i$ is related to the change in surface potential by:

$$\Delta V_i = \Delta V_G - \Delta V_S = \Delta V_G - \left| \frac{kT}{q} \right| (u_s' - u_s)$$  \hspace{1cm} (3.31)

where $\Delta V_S$ is the change in voltage drop over the semiconductor. Combining (3.29)-(3.31) we obtain:

$$\frac{\Delta V_G}{\Delta Q} = \left[ \frac{1}{2q\varepsilon_S N_D} \right] \Delta Q + \frac{1}{C_i} + \frac{\sqrt{2kT/(q^2\varepsilon_S N_D)}}{[u_s]}$$ \hspace{1cm} (3.32)

A plot of $\Delta V_G/\Delta Q$ versus $\Delta Q$ should, therefore, yield a straight line if $N_D$ is constant. $N_D$ can be found easily from the slope of this line.

This method has the advantage that it does not depend on the value of the oxide capacitance and it does not depend on the formation of the inversion layer.

Summarizing Section III.4 it can be said that the most widely used method of finding the doping density at the surface of an oxidized wafer is the high frequency C-V method (Section III.4.1) because it is simple to instrument, fast to analyse (if a computer or desk calculator is available) and can be done on MIS capacitors. It cannot give, however, a doping profile and the accuracy is limited. Better accuracy should be obtained by Kuhn's(71) fast ramp technique which is a little more difficult to instrument but otherwise shows the same advantages. Kuhn's method also gives an indication of the doping profile.
Waxman's\textsuperscript{(74)} technique lays somewhere in between. It cannot give a doping profile but it is easier to instrument than Kuhn's method.

For better accuracies and especially if the doping profile has to be known the Schottky diode technique, especially Copeland's CIP technique\textsuperscript{(73)}, is best suited. It involves, however, the removal of the oxide and the formation of the diodes and is quite complicated to instrument.
Chapter IV

THE IGFET INVERSION LAYER

It has been realized that the inversion layer of an IGFET has some "peculiar" properties. It shows, for example, a strange frequency response and the mobility of the carriers in the inversion layer does not obey the theoretical predictions\(^{76}\). In this chapter some of the properties of the IGFET inversion layer will be investigated. The mobility of the carriers will be covered in Chapter V.

The fact that the inversion layer is quite thin leads to the idea that classical statistics might not be valid to describe the carrier density in the channel. In fact, a quantization of energy levels in inversion layers of metal-oxide-silicon structures has been anticipated for many years\(^{77}-^{82}\). It was, however, believed that quantization effects get washed out at room temperature due to the random scattering of the carriers at the surface. It will be shown in this chapter\(^*\) that for determining the carrier distribution in the inversion layer, quantization effects cannot be ignored even at room temperature. A new definition for the channel thickness will then be introduced and applied to the quantum mechanical and the classical description of the channel. A comparison with other definitions used in the literature will also be made.

Stern and Howard\(^{81}\) treated the surface inversion layer quantum mechanically but restricted themselves to low temperatures. They

\(^{*}\) See also Gnädinger and Talley\(^{83}\).
introduced the notion of electric subbands within the conduction band where every subband corresponds to a quantized level for the motion of the carriers in the direction normal to the surface with a continuum of states for motion in the plane parallel to the surface. Restriction to low temperatures would mean that only the lowest electric subband is occupied. Stern and Howard were able then to obtain the lowest lying energy level and the corresponding charge distribution by a variational principle. This approach will be abandoned in this work since an attempt is made to find the charge distribution in the inversion layer for room temperature where many electric subbands will be occupied and have to be considered.

An exact solution of the problem is mathematically rather involved since—as shown by Stern and Howard—Schrödinger's equation for the motion of the carriers normal to the surface has to be solved self-consistently with Poisson's equation. Schrödinger's equation and Poisson's equation can, however, be decoupled by approximating the exact electrostatic potential by a linear potential. This approximation will be used throughout this chapter. It is then possible to find analytical solutions for the carrier wave functions in the inversion layer and the carrier distribution which will give better physical insight into the behavior of the carriers. As a further simplification, spherical constant energy surfaces, i.e., an isotropic effective mass will be assumed. Also it is assumed that the dispersion relation for the energy is quadratic. These approximations are quite good for electrons in silicon (n-channel devices). Due to the complicated shape of the valence band Coleman \(^{(82)}\) showed that it is not possible to apply the results to p-channel devices. The quantum mechanical calculation
is, therefore, restricted to electrons (n-channel devices). However, the section about channel thickness and shape of the channel which will follow the quantum mechanical calculation is general enough to be applicable to p-channel devices as well.

All results presented in this chapter will be based on the surface potential. A connection to the gate voltage can easily be made by using the information presented in Chapter III.

Before embarking on the detailed analysis it should be noted that the material presented in this chapter and the results obtained are of more academic than practical interest. There is no way really to prove by an experiment the presented theory. Also the differences found between the classical and the quantum mechanical description do not influence the device characteristics. However, it is believed that this work contributes to the understanding of the properties of the inversion layer of an IGFET.

The device configuration and the coordinate system used in this chapter are identical to the one presented in Fig. 3.1. Only the notation has to be changed to n-channel devices.

In the next section the carrier distribution within the inversion layer will be calculated quantum mechanically assuming a constant electric field.

IV.1 Carrier Distribution: Quantum Mechanical Model

Using the coordinate system introduced in Fig. 3.1 we can find the quantized energy levels associated with the motion of the carriers in the x-direction and the corresponding particle wave functions by
solving Schrödinger's equation with an electron potential $U(x)$ given by:

$$U(x) = q \xi x \quad \text{if } x > 0 \quad (4.1a)$$
$$U(x) = \infty \quad \text{if } x < 0 \quad (4.1b)$$

where $\xi$ is the electric field in the semiconductor and $q$ is the electronic charge. As mentioned before $\xi$ is assumed to be constant. The potential energy $U(x)$ is arbitrarily set equal to zero at the interface. Inside the insulator the potential energy is assumed to be infinite since the conduction band in the insulator is much higher in energy than the one in the semiconductor. The quantized energy levels for electrons in a surface channel represented by a triangular potential well have been calculated by Coleman et al.\(^{(82)}\):

$$E_i = -\frac{(q \hbar \xi)^{2/3}}{(2m^*)^{1/3}} x s_i \quad (4.2)$$

where $m^*$ is the effective electronic mass which we shall choose to be equal to $m_0$ and $s_i$ is the $i^{th}$ solution of:

$$Ai(s) = 0 \quad (4.3)$$

$Ai(s)$ is the Airy function. An excellent summary of the properties of Airy functions is given by Abramovitz and Stegun\(^{(84)}\). The first
ten solutions of (4.3) are given in Table I. Higher order solutions can well be approximated by (82):

$$s_i = -\left[\frac{3\pi}{2}(i + \frac{3}{4})\right]^{2/3}. \quad (4.4)$$

The wave function $\Psi_i(x)$ corresponding to the $i^{th}$ quantized energy level is (85):

$$\Psi_i(x) = c_i Ai[(x - E_i/qE)(2m^*qE/h^2)^{1/3}] \quad (x \geq 0) \quad (4.5a)$$

$$\Psi_i(x) = 0 \quad (x < 0) \quad (4.5b)$$

$c_i$ is a constant which can be determined by the normalization condition:

$$|c_i|^2 = \left[\int_0^\infty \left[Ai\{x - E_i/qE)(2m^*qE/h^2)^{1/3}\}\right]^2 dx\right]^{-1} \quad (4.6)$$

As an illustration, in Fig. 4.1 is shown a plot of the energy levels calculated from (4.2) using a doping density of $N_A = 4 \times 10^{15}$ cm$^{-3}$ and two different fields, $3.5 \times 10^4$ V/cm and $1.25 \times 10^5$ V/cm, corresponding to weak and strong inversion. Fig. 4.2 shows a graph of two wave functions corresponding to the lowest and the 7th energy level respectively. The electric field chosen is the one belonging to strong inversion.

In order to be able to calculate the carrier density in the surface inversion layer the density of states function must be known. In a potential well where there exists a discrete set of energy levels
TABLE I: First ten solutions of $Ai(s) = 0$.

<table>
<thead>
<tr>
<th>$s_1$</th>
<th>$-2.3381$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_2$</td>
<td>$-4.0879$</td>
</tr>
<tr>
<td>$s_3$</td>
<td>$-5.5205$</td>
</tr>
<tr>
<td>$s_4$</td>
<td>$-6.7867$</td>
</tr>
<tr>
<td>$s_5$</td>
<td>$-7.9441$</td>
</tr>
<tr>
<td>$s_6$</td>
<td>$-9.0226$</td>
</tr>
<tr>
<td>$s_7$</td>
<td>$-10.0402$</td>
</tr>
<tr>
<td>$s_8$</td>
<td>$-11.0086$</td>
</tr>
<tr>
<td>$s_9$</td>
<td>$-11.9360$</td>
</tr>
<tr>
<td>$s_{10}$</td>
<td>$-12.8288$</td>
</tr>
</tbody>
</table>
Fig. 4.1 Quantized energy levels $E_i$ in a linear potential well at the surface of Si for a doping density of $4 \times 10^{15}$ cm$^{-3}$ and

a) weak inversion ($u_S = 2.57 u_F$)

b) strong inversion ($u_S = 2.71 u_F$)

Fig. 4.2 Normalized wave functions belonging to
a) ground state,
b) 7th quantized level for a linear potential well at the surface of a Si device with a doping density of $4 \times 10^{15}$ cm$^{-3}$ and strong inversion ($u_S = 2.71 u_F$).
associated with the motion perpendicular to the interface and a continuum of states in the planes parallel to the interface the dispersion relation for electrons can be written as \(^{(77)}\):

\[
E = \left(\frac{\hbar^2}{2m^*}\right)(k_y^2 + k_z^2) + E_i
\]

We shall again assume \(m^*\) and \(m_0\) to be equal. \(k_y\) and \(k_z\) are the \(y\) and \(z\) components of the wave vector \(k\). The effective density of states is given by the product

\[
D(E) = D_1(E) D_2(E)
\]

where \(D_1(E)\) is the density of states for the continuum in the \(yz\)-planes and \(D_2(E)\) is the density of states associated with the discrete set of states in the \(z\)-direction. \(D_1(E)\) can be found from

\[
D_1(E) = \frac{1}{A} \frac{d\Gamma(k)/dk}{dE/dk}
\]

where \(A\) is the area of a unit cell in the reciprocal lattice space and \(\Gamma(k)\) is the number of states contained within a circle of radius \(k\) in \(k\)-space:

\[
\Gamma(k) = \pi k^2 A / 4\pi^2 = Ak^2 / 4\pi.
\]

\(E(k)\) is given by the first term of (4.7), i.e.,

\[
E(k) = \left(\frac{\hbar^2}{2m^*}\right)(k_y^2 + k_z^2)
\]
Using (4.10) and (4.11) we can write $D_1(E)$ from equation (4.9) as

$$D_1(E) = \frac{m^*}{2\pi \hbar^2} \quad (4.12)$$

$D_2(E)$ is the product of the number of accessible subbands at the energy considered and the degeneracy of the particular subband per given spin orientation. Consistent with the simplified model chosen in this theory (spherical energy surfaces, isotropic effective mass), the degeneracy of the subbands is taken to be one.\

$D_2(E)$ can be written as:

$$D_2(E) = \sum_i H(E - E_i) \quad (4.13)$$

where $E_i$ is the lowest energy of the $i^{th}$ electric subband and $H(x)$ is defined as

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (4.14)$$

The total density of states is now

$$D(E) = \frac{m^*}{\pi \hbar^2} \sum_i H(E - E_i) \quad (4.15)$$

* If we were to consider the actual band structure of silicon, the degeneracy would depend on the particular surface orientation chosen(75).
where equations (4.8), (4.12) and (4.13) have been used and a factor of two for spin has been included. The total number of electrons in the inversion layer is:

\[ N = \int D(E)f(E)\,dE \]  \hspace{1cm} (4.16)

where \(D(E)\) is the density of states function and \(f(E)\) is the Fermi-Dirac distribution. Inserting (4.15) and the well known form for \(f(E)\), we obtain:

\[ N = \left(\frac{m^*}{\pi\hbar^2}\right) \int \sum_i \frac{H(E-E_i)dE}{1+\exp[(E-E_F)/kT]} \]  \hspace{1cm} (4.17)

where \(E_F\) is the Fermi energy. Because the series in (4.17) converges uniformly we can interchange summation and integration. Using the following new variables:

\[ \xi = \frac{E-E_i}{kT} \hspace{1cm} \eta = \frac{E_F-E_i}{kT} \]  \hspace{1cm} (4.18)

we obtain:

\[ N = \left(\frac{m^*}{\pi\hbar^2}\right)kT \sum_i \int_0^\infty \frac{df}{1+\exp(\xi-\eta)} \]  \hspace{1cm} (4.19)
where the definition of H, eq. (4.14) has been used. The integral in (4.19) is a standard Fermi-Dirac Integral with exponent zero, whose solution is:

\[
F_0(\eta) \equiv \int_0^\infty \frac{df}{1 + \exp(f - \eta)} = \ln[1 + \exp(\eta)] 
\]

(4.20)

Equation (4.19) takes now the form:

\[
N = \left(\frac{m^* kT}{\pi \hbar^2}\right) \sum_i \ln[1 + \exp((E_F - E_i)/kT)] 
\]

(4.21)

where \(\eta\) has been replaced by (4.18). Using the bottom of the conduction band at the surface as reference, the position of the Fermi level \(E_F/kT\) can be expressed in terms of the band parameters as follows (see Fig. 4.3):

\[
E_F/kT = -u_F - (E_g/2kT) + u_s 
\]

(4.22)

where \(E_g\) is the band gap energy and \(u_s\) and \(u_F\) are the dimensionless surface potential and Fermi potential respectively as defined in Chapter III. The carrier density as a function of \(x\) can now be obtained by multiplying the number of carriers in the \(i^{th}\) electric subband by the probability \(p = |\Psi_i(x)|^2\) that a carrier is at a position \(x\) and summing over all occupied subbands. \(\Psi_i(x)\) is the normalized wave function corresponding to the \(i^{th}\) energy level \(E_i\). Thus:
Fig. 4.3 Definition of energy levels and band structure of an inverted p-type semiconductor surface.

Fig. 4.4 Electron density as a function of distance perpendicular to the surface of a Si-device with a doping density of $4 \times 10^{15} \text{ cm}^{-2}$, weak inversion ($u_S = 2.57 \ u_T$)

a) Classical calculation 

b) Quantum mechanical calculation.
\[ n(x) = \sum_i N_i |\Psi_i(x)|^2 \]  

(4.23)

where \( N_1 \) is the number of electrons occupying the \( i \)th electric subband. Using equation (4.21) for \( \sum_i N_i \) we obtain:

\[ n(x) = (m^*kT/\pi\hbar^2) \sum_i \epsilon_n [1 + \exp((E_F - E_i)/kT)] |\Psi_i(x)|^2 \]  

(4.24)

Inserting the explicit forms of the normalized wave functions from equation (4.5) we obtain the final form of the carrier density as a function of distance normal to the interface:

\[ n(x) = (m^*kT/\pi\hbar^2) \sum_{i=1}^{\infty} \epsilon_n [1 + \exp((E_F - E_i)/kT)] |c_i|^2 |A_i[(x-E_i/q\psi)(2m^*q\psi/\hbar^2)^{1/2}]|^2 \]  

(4.25)

where \( |c_1|^2 \) is given by (4.6). It should be recognized that (4.25) is valid only for \( x \geq 0 \), that is in the semiconductor. \( E_F \) is related to the surface potential \( \psi_s \) and the Fermi potential \( \psi_F \) through (4.22). Thus \( n(x) \) in (4.25) is a function of \( x \) with \( \psi_s \) and \( \psi_F \) as parameters. It has to be noted that the electric field cannot be specified independently of \( \psi_s \) and \( \psi_F \). If \( \psi_s \) and \( \psi_F \) are given then \( \psi \) is fixed. This relation will be discussed in more detail later.

In Figs. 4.4 and 4.5 carrier densities as a function of distance are plotted for a doping density of \( N_A = 4 \times 10^{15} \) cm\(^{-3} \) and two different surface potentials, corresponding to weak and strong inversion.
Fig. 4.5 Electron density as a function of distance perpendicular to the surface of Si-device with a doping density of $4 \times 10^{15} \text{cm}^{-2}$, strong inversion ($u_s = 2.71 \ u_F$)

a) Classical calculation

b) Quantum mechanical calculation.
respectively. These graphs were calculated with the help of a computer. A FORTRAN IV program has been developed for this purpose.* The same program is also used to calculate the quantum mechanical channel thickness which will be treated in a later section.

Going back to Fig. 4.1 it can be seen that the spacing of the discrete energy levels decreases for increasing energy and increases for increasing surface potential (electric field), as expected. The lowest energy level for weak inversion lays approximately 0.02 eV (about 1kT) above the bottom of the surface well whereas the lowest energy level for strong inversion is approximately 0.5 eV above the bottom of the well (greater than kT).

We might ask now how many electric subbands are occupied at room temperature as a function of the surface potential in a practical calculation. These subbands are the ones which must be considered in the summation of equation (4.25). The contribution of the different terms in the sum in (4.25) is determined by the Fermi-Integral

\[ F_\theta = \ln \left[ 1 + \exp \left\{ \frac{(E_F - E_i)}{kT} \right\} \right]. \quad (4.26) \]

The exponent in (4.26) can be written as

\[ \frac{(E_F - E_i)}{kT} = -\mu_F - \left( \frac{E_g}{2kT} \right) + \mu_s - \left( \frac{E_i}{kT} \right) \quad (4.27) \]

* A copy of the program is available on request.
where equation (4.22) has been used. Equation (4.26) is now of the form:

\[ F_0 = \ln[1 + \alpha \exp(-E_i/kT)] \quad (4.28) \]

where

\[ \alpha = \exp[\mu_s - \mu_F - (E_g/2kT)] \quad (4.29) \]

For a particular device and a fixed surface potential \( \alpha \) is a constant. If we now terminate the sum in (4.25) when the Fermi-Integral has dropped to 1% of its value for the first subband, we can find the highest energy level which must be considered by setting:

\[ \frac{F_{\text{max}}}{F_{\text{at}}} = \frac{\ln[1 + \alpha \exp(-E_m/kT)]}{\ln[1 + \alpha \exp(-E_i/kT)]} \leq 0.01 \quad (4.30) \]

where \( E_1 \) is the lowest electric subband and \( E_m \) the highest electric subband which will be considered in the summation. Since \( \exp(-E_1/kT) \ll 1 \) for all practical cases we can expand (4.30) to give

\[ \Delta E = E_m - E_i \geq kT \ln(100) \approx 0.117 \text{ eV} \quad (4.31) \]

for room temperature, independent of the surface potential and the doping density.
Since the spacing of the energy levels increases with increasing surface potential the number of energy levels which have to be considered decreases considerably for increasing surface potential, typical values being 30 for weak inversion and 5 for strong inversion.

IV.2 Carrier Distribution: Classical Model

We will now calculate the carrier densities in the inversion layer using classical statistics and compare them to the results obtained quantum mechanically.

In thermodynamic equilibrium and in the classical limit the electron concentration is given by:

\[ n(x) = n_i \exp[u(x) - U_F] \]  \hspace{1cm} (4.32)

where \( u(x) \) is the electrostatic potential at the position \( x \) (see Fig. 4.3) and \( n_i \) is the intrinsic carrier concentration. \( u(x) \) can be obtained by integrating a one dimensional Poisson's equation twice and applying appropriate boundary conditions. The result of this calculation is(86):

\[ x = -L_D \int \frac{du}{u_s} \frac{du}{F(u, U_F)} \]  \hspace{1cm} (4.33)

where \[ F(u, U_F) = \sqrt{2[\cosh(u-U_F) - \cosh(U_F) + u \sinh(U_F)]^2} \]  \hspace{1cm} (4.34)
and $L_D$ is the intrinsic Debye length given by (3.14). An inversion of eq. (4.33) gives the desired $u(x)$. Substituting $u(x)$ into (4.32) gives then the carrier distribution as a function of distance normal to the interface. This analysis can only be carried out numerically. A computer program (FORTRAN IV) has been written to that purpose.*

In Figs. 4.4 and 4.5 the results of such a calculation are shown where the same parameters were used as for the quantum mechanical calculation. A comparison shows that the carrier density calculated quantum mechanically deviates considerably from the classical distribution in the vicinity of the interface for both inversion conditions. This comes mainly from forcing the charge density to be zero at the interface. For distances greater than approximately 40 Å the carrier density approaches the classical limit closely for weak inversion condition (Fig. 4.4). This seems to be reasonable since for weak inversion many electric subbands are occupied so that quantum effects get washed out. For strong inversion on the other hand, as shown before only very few electric subbands are occupied even at room temperature so that quantum effects are important and the carrier density calculated quantum mechanically deviates considerably from the classical limit. (Fig. 4.5)

As Kamin and Muller already pointed out(87) the distribution of the carriers depends critically on the particular statistics chosen whereas the integrated charge in the channel is nearly independent of the statistics. The same statement applies to the comparison of the

---

* A complete listing of the program is available on request.
quantum mechanical description of the channel as presented in this chapter and the classical picture. The carrier density as shown in Figs. 4.4 and 4.5 depends strongly on the way it is calculated whereas the integrated charge (total number of carriers per unit area in the channel) is nearly independent of the method chosen. Quantities which depend only on the total charge in the channel such as conductance or drain current can, therefore, be derived from classical statistics.

It should be noted that the absolute value of the carrier density and, therefore, the integrated charge in the channel depends upon the relation between Fermi level, surface potential and electric field. (See equation (4.25)). The Fermi level and the surface potential can be chosen independently whereas the electric field \( \mathcal{E} \) (assumed to be constant) depends on \( u_s \) and \( u_F \) and should represent a reasonable average over the channel thickness. In the numerical calculations which led to the plots in Figs. 4.4 and 4.5, \( u_s \) and \( u_F \) were chosen independently and \( \mathcal{E} \) was taken to be half of the value it would have at the surface if it were calculated classically from \( u_s \) and \( u_F \)\(^{24}\).

IV. 3 Channel Thickness

We will now turn our attention to a quantity which has received practically no attention in the past: the channel thickness\(^{88}\).

There is no unique way of defining the thickness of an inversion layer and there exist, therefore, various definitions in the literature. The channel thickness is often defined as the distance \( x_i \) perpendicular to the semiconductor - insulator interface, where the semiconductor is just intrinsic (see Fig. 4.3)\(^{78},(87),(89)\). For a
given semiconductor, temperature and doping density, \( x_1 \) is a unique function of the electric field \( E_{xs} \) normal to the surface at the surface. Using classical statistics as described before, \( x_1 \) can be obtained from the condition:

\[
U(x_1) = u_F
\]

(4.35)

Substituting (4.35) into (4.33) gives \( x_1 \) as a function of doping density and surface potential for the classical model:

\[
x_1 = -L_D \int \frac{d\phi}{\phi} - \frac{u_F}{F(u_s, u_F)}
\]

(4.36)

\( F(u_s, u_F) \) is given by (4.34). A family of \( x_1(u_s) \) curves with doping density \( N_A \) as parameter is shown in Fig. 4.6. These curves were obtained by solving (4.36) numerically. \( x_1 \) is not defined for surface potentials \( u_s \) smaller than \( u_F \). For \( |u_s| \geq |u_F| \), \( x_1 \) increases monotonically with increasing \( u_s \), the increase being strongest for smallest \( u_F \) (low doping density). The absolute values attained by \( x_1 \) in the operating range of the IGFET (\( u_s \approx 2u_F \)) decrease strongly with increasing doping density, being approximately 6000 \( \AA \) for \( N_A = 10^{14} \) cm\(^{-3} \) and 400 \( \AA \) for \( N_A = 10^{17} \) cm\(^{-3} \). In Fig. 4.7 \( x_1 \) is shown as a function of the surface field \( E_{xs} \). The doping density in this example is \( 4.6 \times 10^{14} \) cm\(^{-3} \). \( u_s \) is related to \( E_{xs} \) by:

\[
E_{xs} = \frac{2\pi \eta l_D}{\varepsilon_s} F(u_s, u_F)
\]

(4.37)
Fig. 4.6  Intrinsic distance $x_i$ as a function of surface potential $u_s$ with doping density $N_A$ as parameter.
Fig. 4.7 $x_i$ and channel thickness $\delta_c$ as a function of normal electric field $E_{xs}$ at the surface.
where \( \varepsilon_s \) is the dielectric constant of silicon. \( x_i \) as a function of \( E_{xs} \) first increases strongly but levels off at a value of \( E_{xs} \) corresponding to the onset of strong inversion (\( u_s \approx 2u_F \)). \( x_i \) is not defined for surface fields smaller than the "pinch-off" field \( E_p \), where \( E_p \) corresponds to a surface potential \( u_s \) equal to the Fermi potential \( u_F \). \( E_p \) is easily obtained by substituting \( u_F \) for \( u_s \) in (4.37):

\[
E_p = \left( \frac{2\pi n_i L_D/\varepsilon_s}{F(u_F, u_F)} \right)
\]

(4.38)

For a doping density of \( 4.6 \times 10^{14} \) cm\(^{-3} \), \( E_p = 6.2 \times 10^3 \) V/cm.

It will now be shown that \( x_i \) is not a useful definition of channel thickness. A reasonable definition of the channel thickness has to be based on the total mobile charge in the channel\(^{81}\). The following definition is proposed here and will be used in this thesis:

\[
\int_0^\infty [n(x) - n_0] \, dx = 0.99 \int_0^\infty [n(x) - n_0] \, dx
\]

(4.39)

\( n_0 \) is the equilibrium electron density in the bulk and \( n(x) \) is the electron density as a function of \( x \). The channel thickness \( \delta_c \) is thus defined as the distance perpendicular to the interface in which 99% of the total electron charge per unit area is located. This definition applies to any functional dependence of the carrier density. Definition (4.39) will now be applied to the classical model of the channel. The carrier distribution, equation (4.32), is inserted into (4.39) and the integrals evaluated numerically. A family of \( \delta_c \) curves as function of surface potential with doping density as parameter is shown in Fig. 4.8.
Fig. 4.8: Channel thickness $\delta_c$ as a function of surface potential with doping density $N_A$ as parameter. The dashed line corresponds to the onset of strong inversion, $\delta_c(2u_F)$. 

$N_A, \,(cm^{-3})$

$1 \times 10^{14}$

$5 \times 10^{14}$

$1 \times 10^{15}$

$5 \times 10^{15}$

$1 \times 10^{16}$

$5 \times 10^{16}$

$1 \times 10^{17}$

$\delta_c, \,(\AA)$

$U_S - U_F$
In Fig. 4.7 channel thickness $\delta_c$ is shown as a function of surface field $E_{xS}$ for $N_A = 4.6 \times 10^{14}$ cm$^{-3}$. As can be seen clearly $\delta_c$ and $x_1$ deviate drastically. $\delta_c$ decreases for increasing doping density. $\delta_c$ also decreases for increasing surface potential (surface field). This is physically reasonable since the carriers are—in the average—closer to the surface for stronger fields. For a doping density of $4.6 \times 10^{14}$ cm$^{-3}$ $\delta_c$ is approximately 800 Å in the operating range of the IGFET ($E_{xS} \sim 10^5$ V/cm), nearly an order of magnitude smaller than $x_1$. This discrepancy comes from the fact that the carrier density has to drop over many orders of magnitudes and may trail off quite slowly before reaching the intrinsic point, whereas the main part of the carriers is confined to a rather narrow region at the surface. $x_1$, therefore, overestimates the channel thickness considerably and does not represent a reasonable definition of channel thickness.

The new definition of channel thickness, equation (4.39) will now be applied to the quantum-mechanical model of the channel. The carrier distribution (4.25) is inserted into (4.39) and the analysis performed as for the classical case. The resultant channel thickness as a function of surface potential is shown as curve (a) in Fig. 4.9 for a doping density of $4 \times 10^{15}$ cm$^{-3}$. Curve (b) in the same graph is based on the same channel thickness definition (4.39) but using the classical carrier distribution as described before.

A different definition of channel thickness has been proposed by Schrieffer(77). He defines the channel thickness $\delta_c$ as the classical turning point of free carriers of thermal energy in the surface potential well and obtains for a linear potential:
\[ \delta_c = \frac{kT}{xq\xi} \]  

(4.40)

where \( \xi \) is the electric field in the space charge region assumed to be constant.

Using this definition the resultant channel thickness versus surface potential curve is shown as curve (c) in Fig. 4.9. Finally curve (d) gives the distance \( x_1 \) from the interface where the semiconductor is just intrinsic.

The channel thickness as defined in equation (4.39) is derived from the integrated charge. Its value and its dependence on the surface potential should, therefore, not depend considerably on whether classical or quantum mechanical calculations are used. Curves (a) and (b) in Fig. 4.9 bear out this fact. Curve (a) (quantum mechanical) shows a somewhat weaker dependence on surface potential than curve (b) (classical). This can be explained by the fact that increasing the surface field shifts the quantized levels for electrons up rather than narrowing the spatial distribution of the carriers\(^\text{78}\) as can be seen from Figs. 4.1, 4.4, and 4.5. The quantum mechanical channel thickness (curve (b)) decreases monotonically. Curve (c) in Fig. 4.9 based on Schrieffer's definition (4.40), however, shows much lower values of the channel thickness and shows a stronger dependence on surface potential than curves (a) and (b). \( x_1 \), (curve (d)), as mentioned before, overestimates the channel thickness considerably and shows a slight increase of channel thickness with increasing surface potential contradictory to the other definitions.

The analysis and the results presented so far in this chapter are valid for a uniform channel, i.e., the surface potential along the
Fig. 4.9 Channel thickness as a function of surface potential for a Si-device with a doping density of $4 \times 10^{15}\text{cm}^{-3}$.

a) Quantum mechanical, definition equation (4.39)
b) Classical, definition equation (4.39)
c) Definition after Schrieffer equation (4.40)
d) Distance $x$, where semiconductor is just intrinsic, equation (4.36).
channel has been assumed to be constant. This corresponds to a zero drain voltage for the IGFET. We will now look at the channel thickness \( \delta_c \) and the distance \( x_i \) where the semiconductor is just intrinsic for an IGFET operating in the saturation mode (see Chapter II).

The electric field normal to the semiconductor surface, \( E_{xs} \), can now vary substantially along the channel. As shown by Pao and Sah (26) \( E_{xs} \) decreases from source to drain so that \( x_i \) also decreases from source to drain.* In Fig. 4.10a \( E_{xs} \) and \( x_i \) are shown as functions of normalized distance along the channel of an IGFET operating just in saturation. \( E_{xs} \) has been taken from Fig. 10 of reference (26) and was calculated for a particular sample using nonequilibrium analysis and including the effect of diffusion currents. It can be seen that over most of the channel \( x_i \) decreases slightly from source to drain falling off sharply only in the neighborhood of the "pinch-off point" \( y_p \), where it is zero. \( y_p \) is the distance along the channel where \( E_{xs} = E_p \).

For the example in Fig. 4.10 \( y_p/L = 0.922 \).

For an IGFET operated in saturation where the electric field decreases from source to drain the channel thickness \( \delta_c \) as defined in (4.39) increases from source to drain as shown in Fig. 4.11a. Beyond the "pinch-off point" \( y_p \), the channel thickness is not defined since the drain region is essentially depleted of holes and the integral in equation (4.39) becomes meaningless. The "shape" of the channel

---

* Strictly speaking equations (4.32)-(4.38) are only valid in thermal equilibrium, a condition certainly not met by an IGFET operating in saturation. If \( E_{xs}(y) \) is known, however, \( x_i(y) \) can be calculated from equation (4.32)-(4.38) by interpreting \( u_s \) as the surface potential which would exist if the IGFET were in thermal equilibrium having a surface field \( E_{xs}(y) \).
Fig. 4.10 (a) $x_i$, channel thickness, $\delta_c$, and normal field at the surface, $E_{xs}$, as functions of normalized distance along the channel. (b) p-channel IGFET structure showing the shape of the channel.
is then as shown in Fig. 4.10b. (Note that the dimensions are not drawn to scale.)

Summarizing the main features of the discussion of channel thickness: $\delta_c$ decreases with increasing surface field and increases from source to drain—contrary to the behavior of $x_1$.

As mentioned before the discrepancy between $x_1$ and the channel thickness $\delta_c$ has no direct implications on the theory of an IGFET since the macroscopic properties of the transistor (drain current, conductance, etc.) depend on the total charge in the channel and not on the local distribution of the carriers within the channel. One important consequence of this chapter, apart from the concept that the channel widens toward the drain, is perhaps the result that the channel is several hundred Angstroms thick. Therefore, assumptions in the analysis of the IGFET involving the channel thickness have to be made with caution. For example, quite often it is assumed that the oxide thickness of an MIS structure is large compared to the channel thickness. If an IGFET is now operated in weak inversion (near threshold), the channel thickness is roughly 400 Å depending on the surface potential and substrate doping and is comparable to a standard oxide thickness of say 1000 Å. The above assumption is, therefore, not strictly true.

One question remains to be answered: is the quantum mechanical model used valid?

An estimate of the error introduced by not solving Schrödinger's equation and Poisson's equation self-consistently could be made by inserting the mobile carrier density (equation (4.25)) together with the
fixed bulk charge (derivable from $u_F$) into Poisson's equation and integrating once. The electric field obtained by this procedure—averaged over the channel thickness—could then be compared to the originally assumed constant field $\mathcal{E}$. By making this comparison and estimating the influence of any deviations on the carrier density it can be shown that the peak value of the mobile carrier density (Fig. 4.4 and 4.5) and the channel thickness (Fig. 4.7) are slightly underestimated for weak inversion and slightly overestimated for strong inversion by assuming $\mathcal{E}$ to be half the value of the surface field. For the surface potential range considered in this paper the deviation should, however, be not more than 10 percent.
Chapter V

SURFACE MOBILITY NEAR THE TRANSITION REGION

As already mentioned in Chapter I the characteristics of the IGFET in the transition region between depletion and inversion (near threshold) have received little attention in the past\(^\text{20}\),\(^\text{90}\). In this chapter the surface mobility of the carriers in the inversion layer will be investigated with particular reference to the transition region of the IGFET.

First the existing surface mobility theories will be reviewed and compared to each other. Then the experimental techniques of determining surface mobility will be discussed and a new measurement technique will be introduced which allows us to find surface mobility in or near the transition region. Experimental results obtained on standard double insulator p-channel structures (see Chapter II) will be given and compared to the most reasonable surface mobility theory.

V.1 Theories of Surface Mobility

No attempt will be made to cover the various theories of surface mobility in detail. We will only look at the basic assumptions made in the different theories, compare the results obtained and pick out the theory which seems to be the most reasonable. This theory will then be compared to experimental results obtained in this work. The emphasis in this chapter is clearly put on the experimental part and not the theory.
The basic definition of mobility is

$$\vec{v} = \|\mu\| \vec{E}$$

(5.1)

where $\vec{v}$ is the drift velocity under an applied field $\vec{E}$. $\|\mu\|$ is in general a tensor. It is determined by the type and strength of scattering mechanisms a carrier (hole or electron) encounters.

In the bulk of silicon the mobility $\mu_b$ can be represented as

$$\mu_b = \frac{q\tau_b}{m^*}$$

(5.2)

where $m^*$ is a suitably defined effective mass and $\tau_b$ is a relaxation time which depends on various parameters such as energy, momentum, impurity density, etc. and considers the scattering mechanisms a carrier encounters in the bulk. Of course, (5.2) is already a gross simplification, since, strictly speaking, a relaxation time can only be defined for elastic scattering mechanisms$^{(91)}$ and many of the scattering mechanisms in semiconductors are inelastic. Even if we assume that $\tau_b$ can be used, it is very difficult to calculate theoretically. We will not be concerned with bulk mobility in this work and whenever it is needed the appropriate values will be taken from the literature$^{(92),(93)}$.

A theory of surface mobility has now to take into account any scattering mechanisms a carrier encounters in a surface channel in addition to the bulk scattering. Those additional scattering mechanisms are, for example:

1) The scattering of the carriers at the gross discontinuity which is introduced into the crystal lattice by the surface.
This scattering can be either diffuse (component of the
velocity parallel to the surface is randomized) or specu-
lar (parallel velocity component is preserved), as will
be discussed later.

2) Impurity scattering at the charged interface states.

Schrieffer\(^{21}\) was the first to derive a theory of surface mo-
bility considering the first of the above mechanisms. He solved Boltz-
mann's transport equation assuming:

a) Random (diffuse) scattering at the surface
b) Constant relaxation time
c) Constant electric field normal to the surface in the
   space charge region
d) Constant scalar effective mass

Schrieffer treated only the case of carriers attracted to the surface
which is the only case of interest for an IGFET structure. Under the
above assumptions he found the following equation for the surface mo-
bility:

\[
\frac{\mu_{\text{eff}}}{\mu_b} = 1 - \exp(\alpha^2)(1 - \text{erf}\alpha)
\]  

where \( \mu_{\text{eff}} \) is the effective surface mobility, \( \mu_b \) the bulk mobility
and \( \alpha \) is given by

\[
\alpha = (q E_{xs} \tau)^{-1/2}(2m^*kT)
\]

In (5.4) \( E_{xs} \) is the electric field at the surface, \( \tau \) is the relaxation
time of the carriers in the surface well, assumed to be equal to the
bulk relaxation time \( \tau_b \) and \( m^* \) is the effective mass of the carriers.

In Fig. 5.1 \( \mu_{\text{eff}} \) determined from (5.3) is shown as a function of the
parameter \( \alpha \). For an infinite electric field normal to the surface \( \alpha \)
Fig. 5.1 Effective surface mobility for electrons as a function of the parameter $\alpha$, defined in eq. (5.4). (After Schrieffer(21)).

Fig. 5.2 Surface mobility for holes with $J_B = 18.3$ (light holes in Ge at 300 K) (after Greene, Frankl, and Zemel(22)).
would be zero and the surface mobility is zero. For decreasing sur-
face fields, $\alpha$ increases and the surface mobility increases and ap-
proaches asymptotically the bulk mobility for large $\alpha$ (small fields).

There is no way really to prove whether the scattering of the
carriers at the surface of the semiconductor is diffuse or specular.
In analogy to optics one would intuitively assume that the scattering
is partly diffuse and partly specular depending on the angle of inci-
dence and the wave length of the carriers. The probability $p$ that
a carrier is diffusely scattered influences the surface mobility. For
completely specular scattering at the surface the surface mobility is
equal to the bulk mobility since the component of the carrier velocity
parallel to the surface is conserved during a scattering event. For
completely diffuse scattering, however, the surface mobility is re-
duced since the parallel component of the velocity is randomized after
a scattering event. For a mixture of diffuse and specular scattering
the surface mobility, therefore, decreases with increasing diffusivity.
There is, thus, always the temptation to adjust the probability $p$ that
a carrier is diffusely reflected in such a way that the experimental
data of the quantity investigated—in our case the surface mobility—
matches best the existing theory. This approach is, however, not very
sound, since many other parameters can influence the experimental data
as will be shown later. Nevertheless, this approach is taken by many
investigators.

Greene, Frankl and Zemel(22) extended Schrieffer's model by
allowing for partly specular scattering at the surface. They also in-
cluded the case that carriers are repelled from the surface. This has
no great influence on the IGFET inversion layer since one type of carriers, the one which is attracted to the surface, always dominates.

The most important improvement made by Greene et al. is the use of the exact electric field instead of a constant field in the space charge region. The result of this theory is shown in Fig. 5.2 where surface mobility is plotted as a function of surface potential with Fermi potential as parameter for the case of light holes in Ge at 300 \( ^\circ \)K. Completely diffuse scattering was assumed. The following features are immediately evident:

1) For a deep well (\( u_s \ll 0 \) in the present case) the mobility is strongly reduced and agrees well with Schrieffer's approximation in this limit.

2) The mobility is still somewhat reduced for \( u_s = 0 \) (zero surface field) in contrast to Schrieffer's theory.

3) If the carriers are strongly repelled from the surface (\( u_s \gg 0 \) in the present case) the mobility curves approach the bulk value of the mobility. This case was not covered by Schrieffer.

Ham and Mattis\(^{94}\) removed the restriction on a constant effective mass and found deviations up to 20% for Ge from Schrieffer's theory if certain crystal planes were taken as the surface. Greene\(^{95}\) improved the boundary condition of diffuse scattering at the surface which was mentioned above by treating the diffusivity (probability that a carrier is diffusely scattered) as a function of incoming and outgoing direction angles.

Summarizing the theories presented so far which are concerned with the scattering at the surface proper we find the common feature that the surface mobility is reduced for strong attractive electric fields normal to the surface and increases for decreasing electric fields. In the most simple theory by Schrieffer\(^{21}\) the surface mobility approaches the bulk mobility for zero electric field whereas in
the more sophisticated theory(22) this happens only for strong repelling electric fields.

The second scattering mechanism mentioned at the beginning of this chapter—the impurity scattering at the charged interface states—has been treated by Greene and O'Donnell(96), Kamins and MacDonald(97) and Arnold and Abowitz (98) among others. Greene and O'Donnell derived a formula for the diffusivity due to scattering by localized surface charges of a carrier striking the surface, i.e., they considered the influence of a charge center at the semiconductor surface on the probability p that a carrier is diffusely reflected. Kamins and MacDonald(97) compared Greene and O'Donnell's (96) theoretical expression with experimental data. They measured the surface mobility as a function of the gate voltage on MIS structures using Leistiko's(76) method, fitted the data with an empirical relation including the diffusivity parameter p and derived a relation for the diffusivity change Δp as a function of gate voltage V_G. They attributed, therefore, the total change in surface mobility to a change in diffusivity. The experimental values Δp were then compared to the corresponding theoretical values obtained by Greene and O'Donnell. This is illustrated in Fig. 5.3. The match obtained between theory and experiment was quite good. It has to be noted, however, that the concentration of surface charges was made very high, of the order of 10^{12} charge centers per cm^2 by irradiating the devices with electrons. The scattering of the carriers in the surface channel by charged interface states was, therefore, the dominant surface scattering mechanism.

Summarizing the results obtained by Greene and O'Donnell and Kamins and MacDonald we see that for a fixed number of surface scatter-
Fig. 5.3 a) Experimental curves of \((1-p_1)\)
and \((1-p_2)\) as function of \(V_G^*\).
\(p_1, p_2 = \text{diffusivity before and after irradiation of the devices, respectively.}\)

b) \(\Delta p\) as a function of \(V_G^*\).
Solid curve: Theory
Points: Experiment
(After Kamins and McDonald(97)).

Fig. 5.4 Dependence of electron mobility on the number of filled surface states \((Q^* / q)\)
for samples oriented (111) direction (open circles) and (100)(black circles).
The solid line is calculated for mixed lattice and impurity scattering.
(After Arnold and Abowitz(98)).
ing centers the diffusivity decreases with increasing gate voltage (surface potential) resulting in a slower decrease of the surface mobility with increasing gate voltage than expected from Schrieffer's\(^\text{(21)}\) surface mobility theory.

Arnold and Abowitz\(^\text{(98)}\) investigated the change of surface mobility with a change in interface charge. They found a decrease of mobility with increasing interface charge as shown in Fig. 5.4. They explain this behavior by converting the charge in interface states to an equivalent volume concentration and calculating the contribution to the mobility using the conventional theory for impurity scattering in the bulk\(^\text{(99)}\). The solid line in Fig. 5.4 is a theoretical mobility versus interface charge curve obtained by mixing this interface charge contribution with an appropriate lattice scattering part.

Recently Pierret and Sah\(^\text{(23)}\) published a surface mobility theory with particular emphasis on the MIS system. They extended the isotropic model proposed by Schrieffer\(^\text{(21)}\) by using the exact-field analysis and assuming an elliptical energy band structure appropriate for the conduction band of silicon. They found that the anisotropy effect due to nonspherical energy surfaces is less than 10\% (which is to be compared to Ham and Mattis\(^\text{(94)}\) 20\% for Ge) so that nearly always an isotropic model can be used. They also assumed completely diffuse scattering at the surface. Their theoretical expression for the surface mobility as a function of doping density, surface potential and structural parameters is: (p-channel devices)

\[
\frac{\mu_{\text{eff}}}{\mu_b} = 1 - \frac{\rho_b}{\Delta P} \int_{u_s}^{\infty} e^{\nu} (1 - e^{-2k\nu}) \, d\nu \\
\quad (u_s < 0) \tag{5.5}
\]
with \[ \Delta \rho = \hat{u}_s n_i \exp(u_F) \int_{0}^{-u_s} \frac{\exp(u) - 1}{F(u,u_F)} \, du \]  
(5.6)

\[ K_0 = \frac{L_D}{(2\sqrt{\pi} \lambda_p)} \int_{0}^{-u_s} \frac{du'}{F(u',u_F)\sqrt{u' - u}} \]  
(5.7)

\[ \lambda_p = \tau (kT/2\pi m^*)^{\nu_2} \]  
(5.8)

where \[ \hat{u}_s \] is the sign of \( u_s \), \( F(u,u_F) \) is given by equation (4.34), \( \tau \) is the relaxation time assumed to be constant and equal to \( \tau_b \). \( m^* \) is the scalar effective mass. The numerical values assigned to \( m^* \) by Pierret and Sah were:

\[ m^* = m_e^* = 0.263 \, m_0 \quad \text{for electrons} \]
\[ m^* = m_e = 0.369 \, m_0 \quad \text{for holes} \]

where \( m_0 \) is the free electron mass.

Equations (5.5)-(5.8) are written for holes because they will be used for comparison with experimental hole mobilities. Note that this theory is based on an isotropic model which represents the conduction band of silicon better than the valence band. For holes the effective mass \( m^* \) has to be found by averaging over the heavy and light hole band. For electrons the sign of \( u_s \) and \( u_F \) in (5.5)-(5.8) would have to be reversed. In Fig. 5.5 a family of surface hole mobility versus surface potential curves with \( u_F \) as parameter is shown where (5.5)-(5.8) have been used.

Comparing the various surface mobility theories presented so far it is the feeling of the author that Pierret and Sah's (23) theory is the most reasonable one to compare the experimental results with.
Fig. 5.5 Isotropic exact-field theory for hole effective mobilities applicable to inversion-channel IGFET. (After Pierret and Sah\textsuperscript{(23)}).
It is a compromise between complexity and accuracy. It considers the exact-field distribution which is essential for a reasonable accuracy but uses an isotropic model which simplifies the calculation considerably without introducing too large an error. Furthermore, Pierret and Sah present their results for surface potential intervals corresponding to the operating range of IGFETs (strongly inverted surfaces). For devices with large interface state densities it might be necessary to modify this theory by including the additional impurity scattering caused by charged interface states (96)-(98). This can also be the case if the devices are investigated in the transition or threshold region (see next section).

Very briefly the main features of Pierret and Sah's theory can be summarized as follows:

1) For low surface fields (transition region of IGFETs $\mu_{\text{eff}}$ decreases strongly with increasing doping density.

2) For strong inversion $\mu_{\text{eff}}$ is nearly independent of doping density.

3) $\mu_{\text{eff}}$ decreases monotonically with increasing surface potential.

V.2 Experimental Determination of Surface Mobility

There are essentially two experimental techniques for determining surface mobility: (a) Performing a Hall experiment, (b) from surface conductance.
The Hall method, extensively used in practice \(^{(80),(94),(100)-(103)}\) has the following disadvantages: The Hall mobility \(\mu_H\) is not necessarily equal to the conductivity mobility \(\mu\) which we are only interested in. \(\mu\) and \(\mu_H\) are related by:

\[
\mu = (1/r) \mu_H
\]  
(5.9)

where \(r\) is the Hall ratio given by \(^{(104)}\):

\[
r = \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2}
\]  
(5.10)

\(\tau\) is the relaxation time of the carrier which depends, as mentioned in Section V.1, on energy, momentum, crystal orientation, etc.

The Hall ratio depends, therefore, on the scattering mechanisms involved. It has been calculated for some bulk processes, but no attempt has been made so far to calculate \(r\) for surfaces, including the surface scattering mechanisms. The only exception might be a paper by Zemel \(^{(105)}\). He calculated the Hall ratio for a silicon surface under highly idealizing assumptions such as constant electric field, constant relaxation time, scalar effective mass, etc. He finds \(r = 0.87\). This value is, however, not too reliable.

Hall measurements are usually performed on van der Pauw geometries \(^{(100),(106)}\) with Hall contacts located alongside the channel. This structure deviates considerably from the usual IGFET configuration and it is questionable whether the changed geometry preserves the general features of the IGFET such as the current distribution in the channel. In addition to an uncertainty in the Hall ratio \(r\), we, therefore, introduce another possible error by measuring the Hall mobility on a
different structure.

The conductance method, to be described below, avoids the above limitations because it allows us to extract the conductivity mobility directly from a measured channel conductance of the IGFET under consideration. It also yields better information about anisotropy effects which are more difficult to observe with Hall measurement techniques. In this thesis the conductance method will be used exclusively.

Experimental Method

In this chapter we will again assume p-channel devices. The technique of extracting surface mobility of an IGFET structure from measured channel conductance was first employed by Leistiko\(^{105}\) and subsequently used by others\(^{23},^{80}\). Quite generally the mobility of the free carriers in the channel (holes in our case) can be determined from

\[
\sigma_D = \frac{(W/L)}{\mu_p Q_p} \tag{5.11}
\]

which was already introduced in Chapter II (equation (2.1)). \(\mu_p\) is the hole mobility, \(\sigma_D\) the channel conductance, \(L/W\) the length-to-width ratio of the channel and \(Q_p\) the mobile hole charge per unit area in the channel. The main difficulty in this method is the separation of the product of mobility and mobile carrier density in (5.11). The mobile hole charge \(Q_p\), can be calculated as a function of device and materials parameters and surface potential \(u_s\). The channel conductance \(\sigma_p\), however, is measured as a function of gate voltage \(V_G\). Therefore, a relationship between gate voltage and surface potential has to be
established. Having such an expressing $g_p(V_G)$ can then be transformed to $g_p(u_s)$ or $Q_p(u_s)$ can be transformed to $Q_p(V_G)$. Thus, from equation (5.11) we can find the surface mobility either as:

$$\mu_p(u_s) = (L/W) g_p(u_s)/Q_p(u_s) \tag{5.12a}$$

or

$$\mu_p(V_G) = (L/W) g_p(V_G)/Q_p(V_G) \tag{5.12b}$$

The connection between gate voltage and surface potential is usually done on a theoretical basis as shown later. For the IGFET operating in or near the transition region this relationship cannot be established precisely because some parameters, such as interface state density, are not known accurately enough. In this thesis this difficulty is avoided by determining the surface potential as a function of gate voltage experimentally. Using this experimental transformation (5.12) can easily be solved.

We will now cover the different steps in the analysis in more detail. We shall first show how the mobile charge can be found as a function of $u_s$. Then the method for measuring the channel conductance as a function of gate voltage will be presented and finally the experimental technique which permits relating the surface potential to the gate voltage will be given.

**Mobile Hole Charge**

$Q_p$ as a function of $u_s$ is given by (3.12) which is repeated here for convenience:
Chapman

\[ Q_p(u_s) = \frac{q \pi L \lambda}{u_F} \int_{u_F}^{u_s} \frac{\exp(u_F - u) du}{\sqrt{2[\cosh(u_F - u) - \cosh(u_F) + u \sinh(u_F)]}} \]  

(5.13)

The meaning of the symbols is given in Section III.2 and in the "List of Symbols."

In the literature\(^{(20),(76)}\) approximations are usually introduced to avoid a numerical integration in equation (5.13). These approximations are satisfactory for strongly inverted surfaces but fail if the transistor is operated near the transition region. In this work the exact theory was used and the resulting integrals were solved numerically with the aid of a computer (GE 635).

Channel Conductance

The method of measuring small signal channel conductance has been explained in Section III.2 and an illustrative example of a measured conductance versus gate voltage curve has been shown in Fig. 3.7. No dc bias voltage was applied to the drain. It was found that conductance values corresponding to the transition region were sensitive to the amplitude of the measuring signal. Conductance values measured at 10mV peak-to-peak ac drain voltage were usually up to 3 \( \mu \)mhos higher than those measured at 1mV for gate voltages near threshold as shown in Fig. 5.6. If the measuring signal was of the order of 1mV, however, the channel conductance was nearly independent of the measuring signal so that conductance values measured with 1mV ac drain voltage should be close to the zero drain voltage values even for the transition region of the IGFET. Some mechanisms which could influence the channel
Fig. 5.6 Dependence of channel conductance on ac drain voltage.

Fig. 5.7 Experimental and theoretical surface potential versus gate voltage curves.
conductance as a function of ac drain voltage are discussed in the Appendix. The analysis and the results given in the Appendix will be summarized briefly: First, static channel conductance versus drain voltage characteristics were derived where the drain voltage was allowed to have either sign. For positive drain voltages the forward biased drain-substrate junction had to be taken into account. From this static characteristic the contribution to the measured conductance was calculated assuming a sinusoidal ac voltage symmetrical around zero drain voltage. It was found that for ac drain voltages comparable to kT/q (≈ 25mV at room temperature) no conductance-drain voltage dependence of more than a few percent was found. The large dependence demonstrated in Fig. 5.6 could not be explained by the effects covered in the Appendix.

Relation between gate voltage and surface potential

Having found \( Q_p \) as a function of \( u_s \) and \( g_n \) as a function of \( V_G \) a crucial step in the analysis is establishing a relationship between gate voltage \( V_G \) and surface potential \( u_s \). Usually \( u_s \) as a function of \( V_G \) is found by solving the following equation for \( u_s \):

\[
V_G - (q_{m_s} - Q_s/C_i) = (kT/q)u_s - Q_s(u_s)/C_i
\]  
(5.14)

\( Q_s \) is the total induced charge per unit area in the semiconductor given by:

\[
Q_s(u_s) = 2q\rho n_i l_D \left\{2[\cosh(u_F-u_s) - \cosh(u_F) + u_s \sinh(u_F)] \right\}^{1/2}
\]  
(5.15)
and the symbols in (5.14) and (5.15) were defined earlier. Equation (5.14) is only valid for gate voltages well above threshold and cannot be used for gate voltages corresponding to the transition region of the IGFET for the following reasons:

1) Near the transition region the mobile carrier density is of the order of $10^9 \text{cm}^{-2}$. The density of charged interface states which are neglected in (5.14) is of the same order of magnitude and should, therefore, be included. As mentioned in Section III.1, interface state densities cannot be measured accurately, especially not near the band edges where they are needed in the present problem. For strong inversion, however, the total induced charge in the semiconductor is much larger than the charge in the interface states so that little error is made by neglecting them.

2) $(\phi'_{MS} - Q_{SS}/C_i)$ in equation (5.14) (measurement of which was treated in detail in Section III.2) can be measured within approximately $\pm 0.1V$ as mentioned before. In the transition region where $V_G$ is only slightly larger than $(\phi'_{MS} - Q_{SS}/C_i)$, this accuracy is not satisfactory because the error in $u_s$ due to an error in $(\phi'_{MS} - Q_{SS}/C_i)$ increases to first order exponentially with decreasing $|V_{GF}|$. For strong inversion, however, an error of $\pm 0.1V$ in $(\phi'_{MS} - Q_{SS}/C_i)$ has negligible effect on $u_s$.

Instead of using equation (5.14) the relationship between gate voltage and surface potential was established experimentally in the present work.
As mentioned in Section III.1 while treating the measurement of interface states, the surface potential $u_s$ can be found experimentally up to an additive constant as a function of gate voltage $V_G$ from a thermal equilibrium C-V curve as given in (3.6). This technique\(^{(51)}\)\(^{(56)}\) has been extended to establish the required relationship between surface potential and gate voltage. The only difference was the way the additive constant $\Delta$ was found. In contrast to the methods described in Section III.1, we determined the additive constant $\Delta$ by combining equations (3.6) and (5.14) for any gate voltage $V_G$ corresponding to strong inversion. As said before, the theoretical relationship between gate voltage and surface potential given in (5.14) is quite accurate for strong inversion. We can, therefore, find one particular value of $u_s$ for any gate voltage $V_G$ corresponding to strong inversion by solving (5.14). Inserting $u_s$ and $V_G$ thus found into equation (3.6) we can solve this equation for the constant $\Delta$.

This procedure is more accurate than the method described in Section III.1. No intermediate manipulation of measured data such as converting the measured $C(V_G)$ to a $C(u_s - \Delta)$ curve has to be done. Furthermore, the error introduced by taking data measured on an MIS capacitor to determine the surface potential underneath the gate of an IGFET is reduced because $\Delta$ is essentially determined on the IGFET location. An example of an experimental surface potential versus gate voltage curve is shown as a solid line in Fig. 5.7. In this example $\Delta$ was determined by matching (3.6) and (5.14) for $V_G = -3V$. For comparison a theoretical surface potential versus gate voltage curve for an ideal structure is also shown on the same graph. It can be seen that the two curves deviate considerably in the low inversion and the
dépletion region.

V.3 Experimental Results

The surface mobility of the mobile holes as a function of surface potential \( u_s \) and gate voltage \( V_G \) is shown as solid lines in Fig. 5.8a and 5.8b, respectively. These curves represent a typical example of many devices investigated in the manner described above. For comparison the broken line indicates the surface mobility with interface states not considered. It can be seen clearly that the two curves deviate drastically near the transition region of the IGFET but approach each other for higher gate voltages.

As an indication where the transition region is located a line representing a surface potential equal to twice the Fermi potential indicating the onset of strong inversion is shown in Fig. 5.8a. In Fig. 5.8b the voltage \( V_T \) corresponding to this surface potential is given.

In the example shown in Fig. 5.8, the additive constant \( \Delta \) (see equation (3.6)) was determined by equating (5.14) and (3.6) for \( V_G = -4V \) which corresponds to strong inversion for this particular device. The surface mobility values of the mobile holes in the inversion layer are of the order of the bulk mobility near the transition region \( (u_s \sim 2u_F, V_G \sim V_T) \) as can be seen from Fig. 5.8. For increasing absolute values of surface potential or gate voltage, the surface mobility decreases. This behavior is in general agreement with the theoretical surface mobility as given by Pierret and Sah(23)
Fig. 5.8 a) Surface mobility as function of surface potential.

Solid line (curve A): Surface mobility when interface states are taken into account by determining $u_s$ experimentally.

Broken line (curve B): Surface mobility when interface states are neglected.

Dotted line (curve C): Theoretical surface mobility curve after Pierret and Sah (23).
which is shown in Fig. 5.8a and 5.8b for comparison as dotted lines. Fig. 5.9 gives an illustration of the fit between theory and experiment obtained by Leistiko(76) et al. Some slices showed an initial increase of surface mobility with increasing surface potential (gate voltage) as shown in Fig. 5.10. This increase could be explained by a reduction of the impurity scattering of the mobile carriers on the charged interface states with increasing surface potential due to screening of the scattering centers. (See Section V.1).

V.4 Conclusions

Perhaps the most significant finding in this chapter is the general agreement of the experimental surface mobility with the theoretical curve, especially for biasing conditions corresponding to weak inversion. It is, however, not surprising that there remains disagreement between experiment and theory. The still existing deviations can be attributed to various uncertainties in both the experimental and the theoretical parts. In the theoretical expressions (5.5)-(5.8) it is difficult to ascribe a correct value to the effective mass as mentioned in Section V.1. The experimental surface mobility curve is affected by uncertainties in the channel length. In this work short channel devices (L = 5μ) were used where small errors in measuring L have a strong influence on the calculated surface mobility (see equation (5.12)). However, the fit between theory and experiment reported in this thesis is far better than the one obtained by other investigators.(76),(23)
Fig. 5.9 Effective hole mobilities as measured with Leistiko's method and the theoretical hole mobility curves after Pierret and Sah.

Fig. 5.10 Surface mobility as a function of surface potential. Solid line (curve A): Surface mobility when interface states are taken into account by determining $u_s$ experimentally. Broken line (curve B): Surface mobility when interface states are neglected.
It has been shown that it is essential to consider interface states while establishing the relation between surface potential and gate voltage. The neglecting of interface states was probably the reason for the apparent constancy of the surface mobility at approximately half the bulk value for weak inversion in the work of Leistiko (76).

The goal in this chapter was to find the true value of the hole mobility in the channel. In the device equations, however, it might still be of advantage to use a "reduced" effective mobility value as that given by Leistiko (76). In that case interface states do not enter explicitly in the equation, but rather are absorbed in the effective mobility value. It should be noted, however, that this effective mobility is not a true mobility since it stems from an incorrect separation of the mobility-mobile charge product.
Chapter VI

SUMMARY

In this chapter we will recapitulate what has been presented in the preceding five chapters.

The purpose of Chapter I was to serve as an introduction to the subject and content of the following four chapters. This introduction opened with a brief review of the history of IGFETs, stated the still pending problems in their manufacture and theory, and emphasized the goal of this thesis, namely to contribute theoretically and experimentally to the understanding of some features of IGFETs which have received little attention in the past. Of particular concern was the transition or threshold region.

In Chapter II the experimental IGFET structures used in this thesis and the different processing steps for their fabrication were described. A brief phenomenological description of their principle of operation was also given.

In Chapter III the most important parameters which characterize an IGFET were critically discussed with particular emphasis on the techniques of their measurements. These parameters were surface states density, flatband voltage, threshold voltage and surface doping density. A new measurement technique was proposed for determining \( V_o \) (defined in equation 3.11). Experimental results were presented and compared to conventionally determined values.
In Chapter IV a theoretical description of the IGFET inversion layer was given. Considering quantization of energy in the surface channel the carrier distribution was calculated and compared to classical results. A deviation of the quantum mechanical distribution from the classical distribution for strongly inverted surfaces was found. A new definition of the thickness of the inversion layer was introduced which was based on the total mobile charge per unit area in the channel. Using this definition it could be shown that the channel thickness decreases for increasing surface potential, typical values being 400 Å for weak inversion and 50 Å for strong inversion. It was also shown that the channel thickness increases from source to drain if the IGFET operates in saturation, contrary to the usual description in the literature.

In Chapter V the surface mobility of the mobile carriers in the channel of an IGFET was experimentally determined with particular reference to the threshold or transition region. It could be shown that for this region the standard techniques of determining surface mobility were not valid any more. A new measurement method was introduced, the crucial point of which was that the relationship between gate voltage and surface potential was established experimentally. With this method interface states are automatically included in the analysis. It was also shown that interface states have a pronounced effect on the surface mobility in the threshold region of an IGFET and it is imperative to include them in the analysis. The resultant mobility values agreed remarkably well with the theoretical prediction. It could be shown that the surface mobility approaches the bulk values for the limit of weak inversion.
Suggestion for future work:

It has been realized during the course of this thesis that some of the IGFET parameters are quite difficult to measure. For example, a convenient and accurate technique for measuring the doping density at the surface of a semiconductor does not exist. It would be worthwhile to look for new and different physical effects which depend on this doping density and which could be used for its determination.

It is also suggested to perform an error analysis of the low frequency C-V technique used to measure interface states which is reported in Chapter III. It is known that the error is smallest at mid-gap and increases toward the band edges. It would be valuable to have a quantitative description of these errors in a similar way as Sah et al. (50) have done it for the high frequency C-V method.

Another area of further studies could be an investigation of the influences of a nonuniform doping density on the device characteristics of an IGFET. All IGFET theories presented in the literature assume a constant doping density. A nonconstant doping density, however, introduces an electric field normal to the surface and it should influence the device performance, for example, the frequency response.

It is further suggested to extend the experiments performed in Chapter V and to include large channel devices as well as single insulator structures. It would also be interesting to perform the same analysis on n-channel devices. A thorough study (experimentally and theoretically) of the influence of the measuring drain voltage on the channel conductance for the transition region of an IGFET could be another worthwhile extension of this work.
REFERENCES


36. I. Tamm, Physik. Z. Sowjetunion, 1, 733 (1933).


46. C. Séquin and E. Baldinger, "Interface States in Metal-Oxide-Semiconductor Field Effect Transistors," to be published.


66. W. van Gelder, private communication.


92. S. M. Sze and J. C. Irvin, "Resistivity, Mobility and Impurity Levels in GaAs, Ge, and Si at 300°K," Solid State Electron., 11, 599 (1968).


Appendix

DEPENDENCE OF CHANNEL CONDUCTANCE ON DRAIN VOLTAGE

In this appendix we will investigate theoretically why the measured small signal channel conductance depends on the ac-drain voltage. To be able to do that we have to derive first the dc characteristics of the IGFET for values of $V_D$ near $V_D = 0$. We will, therefore, calculate the dc channel conductance as a function of positive and negative drain voltages with gate voltage as parameter. From these static characteristics the contribution to the measured conductance will be calculated assuming a sinusoidal ac-drain voltage symmetrical around $V_D = 0$.

Several effects or physical mechanisms can be imagined to have an influence on the conductance versus drain voltage characteristics: e.g., the forward biasing of the drain-substrate junction during one-half cycle of the ac signal, the nonuniformity of the channel if the drain voltage and the effective gate voltage are comparable, a voltage dependent series capacitance, a varying channel length, etc. The first two of these effects will now be investigated. We will confine our analysis to p-channel devices.

For p-channel devices negative drain voltages correspond to the operating region of the IGFET. Only this situation is covered in the literature\(^{26}\). For positive drain voltages, the pn-junction formed between drain and substrate will be forward biased. If the voltage between drain and substrate is small (compared to $kT/q \approx 25\text{mV}$...
at room temperature) the current flowing through this junction will be small. Of course, the conductance of the drain-substrate junction will increase with increasing forward bias.

Since the IGFET is a completely symmetrical device we can transform the situation of a positive drain voltage into a normal operating mode with negative drain voltage in the following way:

In Fig. A-1a an IGFET is shown with source and substrate both at zero potential and with a drain potential of $+V_D$. The gate has a potential $-V_G$. This situation exists when the ac-drain voltage swings positive. The drain current flows from drain to source. For positive drain voltages the drain really acts as a source and the source as a drain. If we exchange, therefore, the names "source" and "drain" in Fig. A-1a and draw the IGFET rotated by 180° we obtain Fig. A-1b. Physically the situation shown in Figs. A-1a and A-1b are identical, only the former drain is now called source and the former source called drain. The source in Fig. A-1b is at a potential $+V_D$ and drain and substrate are at zero potential. The current flows from source to drain as in the normal operating mode. The gate is still at a potential $-V_G$. Since a potential is defined only with respect to an arbitrary reference level we can subtract a value $-V_D$ from every potential in Fig. A-1b without changing anything. We then end up with the situation shown in Fig. A-1c. This is the usual connection of an IGFET with the source at zero potential, a negative drain voltage $-V_D$ and a gate voltage ($-V_G - V_D$). The only deviation from a standard connection is the negative voltage applied between substrate and source.

* In this paragraph only $V_G$ and $V_D$ are used as potentials relative to the designated reference point.
Fig. A.1. Transformation of p-channel IGFET with positive drain voltage to normal mode with negative drain voltage.

Fig. A.2 IGFET with negative bias applied to substrate.
This can, of course, give rise to a large conductance \( g_J \) in parallel to the channel conductance if the negative substrate bias is large enough as shown in Fig. A-2. \( g_J \) is the junction conductance between drain and substrate in the actual IGFET configuration (Fig. A-1a). After performing the above transformation for positive drain voltages, \( g_J \) appears between source and substrate (Fig. A-1c and A-2). In the following analysis we will call \( g_J \) the drain to substrate junction capacitance having the actual IGFET configuration in mind.

**Static Drain Conductance as Function of Drain Voltage:**

We will now derive the static drain conductance \( g_D \) (conductance between source and drain with the substrate tied to the source) of an IGFET as a function of drain voltage \( V_D \), where \( V_D \) can be positive or negative, and with the gate voltage \( V_G \) as parameter. For positive values of \( V_D \) the transformation shown in Fig. A-1c is used so that the analysis for negative values of \( V_D \) can be applied. The first two effects mentioned above, namely the nonuniformity of the channel and the forward biased drain to substrate junction will be considered. The total conductance \( g_D \) between source and drain can be written as:

\[
g_D = g_C + g_J \tag{A-1}
\]

where \( g_C \) is the channel conductance and \( g_J \) is the conductance of the drain to substrate junction. (See Fig. A-2).

For negative drain voltages the channel conductance \( g_C \) as a function of \( V_D \) and \( V_G \) has been derived by Pao and Sah\(^{(26)}\) based on a model which includes the diffusion current and is given by (p-channel):
\[ g_C(u_G, u_D) = -2(W/L) \mu_p q \eta L_D \int \frac{e^{u_F - u + u_D}}{F(u, u_D, u_F)} \, du \]
\[ F(u, f, u_F) = \left\{ e^{u - f - u_F} + e^{u_F - u} + (u - 1) e^{u_F} - (u - e^{f} - u_F) e^{u_F} \right\}^{\frac{1}{2}} \]

(A-2)

(A-3)

The upper limit in (A-2) is the surface potential at the drain. \( u_D \) and \( u_G \) are the gate voltage and drain voltage in \( kT/q \) units, respectively. The other symbols were defined in Section III.2 and the List of Symbols. \( u_s(u_D) \) is found from an equation equivalent to (5.14) but including the influence of \( u_D \) (26):

\[ V_G - V_{FB} = (kT/q)u_s - (2q \eta L_D / C_i) F(u_s, u_D, u_F) \]

(A-4)

where \( V_{FB} \) is the flatband voltage and \( F(u_s, u_D, u_F) \) is given by (A-3). The interface states are neglected in the analysis to follow. \( V_{FB} \) is taken to be equal to \( V_o \) (see Section III.2). The results obtained are, therefore, not strictly true for the transition region of an IGFET.

For positive drain voltages we have to reverse the sign of \( u_D \) in (A-2) which then reads

\[ g_C(u_G, -u_D) = -2(W/L) \mu_p q \eta L_D \int \frac{e^{u_F - u - u_D}}{F(u, u_D, u_F)} \, du \]

(A-5)

The upper limit in (A-5) has now to be determined from

\[ V_D + V_G - V_{FB} = (kT/q)u_s - (2q \eta L_D / C_i) F(u_s, u_D, u_F) \]

(A-6)
instead of (A-5) according to the transformation rule given in Fig. A-1.

In Fig. A-3 the channel conductance $g_C$ obtained from equations (A-2) and (A-5) is shown as a function of $u_D$ and $u_G$ for a particular example. The numerical calculations were performed with the aid of a computer.*

Fig. A-3 shows that all curves have their maximum at zero drain voltage. For positive and negative drain voltages the channel conductance decreases, the decrease being strongest for negative drain voltages (normal mode of operation). The change in $g_C$ is more pronounced for low gate voltages. This is easy to understand. For gate voltages near the onset of inversion a very small drain voltage can "pinch-off" the channel and introduce a strong decrease in $g_C$. For gate voltages well above threshold (in strong inversion) the channel conductance $g_C$ is nearly independent of $u_D$ for the values of $u_D$ shown in Fig. A-3.

The conductance in (A-1) of the drain-substrate junction for both positive and negative drain voltages is given by**

$$g_J = \frac{q}{mkT} I_S \exp \left( \frac{u_D}{m} \right)$$  \hspace{1cm} (A-7)

where $I_S$ is the reverse saturation current of the drain-substrate junction which is of the order of nanoamperes in today's devices. $m$ is a constant which equals 1 for ideal current and 2 for generation-recombination current. In the following $m = 1$ is assumed for simplicity.

* A complete listing of the FORTRAN IV program used is available on request.

** See for example Shockley et al. (107).
Fig. A.3  Dependence of channel conductance $g_c$ on drain voltage $u_D$ (kT/q units) for gate voltages $V_G$ belonging to the transition region.
Fig. A-4 shows $g_J$ as a function of $u_D$ for a practical example, shown relative to $g^0_J$ the zero voltage junction conductance.

The total conductance $g_D$ can now be found by inserting (A-7) and either (A-2) or (A-5) into (A-1) depending on whether $u_D$ is positive or negative

$$g_D = -(2W/L)\mu_P q n_p l_p \int_{u_F}^{u_D} \frac{\exp(u_F - u + u_p) du}{F(u, -u_D, u_F)} + \frac{(q/mkT)}{I_s} \exp(u_D/m)$$  \hspace{1cm} (u_D < 0, normal mode)  \hspace{1cm} (A-8)

$$g_D = -(2W/L)\mu_P q n_p l_p \int_{u_F}^{u_D} \frac{\exp(u_F - u - u_F) du}{F(u, u_D, u_F)} + \frac{(q/mkT)}{I_s} \exp(u_D/m)$$  \hspace{1cm} (u_D > 0, reverse mode)

Fig. A-5 shows the conductance $g_D$ as a function of $u_G$ and $u_D$ where the same example as in Fig. A-3 and A-4 has been chosen. Fig. A-5 shows that in the normal mode of operation ($u_D < 0$ for p-channel device) the conductance $g_D$ is always smaller than the zero drain voltage conductance $g^0_D$. For large enough positive drain voltages, however, $g_D$ can become larger than $g^0_D$ and increase strongly with increasing drain voltage due to the influence of the exponentially increasing $g_J$ (see equation (A-7)). This is especially true for gate voltages near threshold.

Measured ac-drain conductance as function of the amplitude of the measuring signal:

The results obtained in the preceding paragraph will now be used to predict theoretically the conductance which would be measured
Fig. A.4  Junction conductance $g_j$ as a function of drain voltage $u_D$ (kT/q units) relative to the zero drain voltage conductance.
Slice ± 03112  FET 3 # 1
N₀ = 9.48 × 10¹⁶ cm⁻³
Vᵣ = -1.02V

\[ g_D / g_{DO} \]

\[ U_D, (kT/q \text{ units}) \]

\[ V_G = -1.7 \]
-1.2
-1.0
-0.9
-0.8

\[ V_G = -0.8 \]
-0.9
-1.0

Fig. A.5 Total conductance \( g_D \) between drain and source relative to the zero drain voltage conductance \( g_{DO} \) as a function of drain voltage \( U_D \) with gate voltages \( V_G \), corresponding to the transition region as parameter.
as a function of the amplitude of the measuring ac-drain voltage with
gate voltage as parameter in the experimental setup shown in the in-
sert to Fig. 3.2. The contribution of $g_C$ and $g_J$ are treated separately.
It is assumed that in the conductance bridge the measured conductance
is determined by a ratio of effective current $i_{\text{eff}}$ and effective vol-
tage $v_{\text{eff}}$:

$$g^m = \frac{i_{\text{eff}}}{v_{\text{eff}}} \quad (A-9)$$

The superscript $m$ stands for measured quantity. The applied ac-drain
signal $v_D$ is sinusoidal and swings symmetrically around $V_D = 0$, i.e.,
no dc-bias was applied to the drain.

$$v_D = V_{D0} \sin(\omega t) \quad (A-10)$$

The effective value of any time varying periodic function $f(t)$ is
defined by

$$f_{\text{eff}} = \left\{ \frac{1}{T'} \int_0^{T'} [f(t)]^2 dt \right\}^{\frac{1}{2}} \quad (A-11)$$

where $T'$ is the time period of the function $f(t)$.

We will now calculate the contribution to the measured con-
ductance from the channel conductance $g_C$. For reasonably small drain
voltages (compared to $kT/q \approx 25\text{mV}$) $g_C$ in equations (A-2) and (A-5) can
be approximated by two straight lines as shown in Fig. A-6.

$$g_C = g_{CO} (1 - \alpha V_D) \quad (V_D \geq 0) \quad (A-12)$$

$$g_C = g_{CO} (1 + \beta V_D) \quad (V_D \leq 0)$$

where $\alpha$ and $\beta$ are the angles defined in Fig. A-6 and are functions
of $V_G$. 

Fig. A.6 Straight line approximation for the channel conductance $g_c$ if $V_D \leq kT/q$ and $V_G$ near threshold.

Fig. A.7 Contribution $g_c^m$ to the measured channel conductance $g_D^m$ due to the nonuniform channel as a function of drain voltage $U_D$ (in units of $kT/q$) with $V_G$ as parameters. The values of $V_G$ correspond to the transition region of the IGFET.
The current flowing due to $V_D$ as a function of $u_D$ is:

$$i(u_D) = \int_{-\infty}^{u_D} g_C(u_D') \, du_D' = g_{CO} \int_{-\infty}^{u_D} (1 - \alpha u_D') \, du_D' \quad (u_D > 0)$$

$$= g_{CO} \int_{-\infty}^{u_D} (1 + \beta u_D') \, du_D' \quad (u_D \leq 0)$$

Performing the integrations in (A-13) we obtain:

$$i(u_D) = g_{CO} u_D - \frac{g_{CO} \alpha}{2} u_D^2 \quad (u_D > 0)$$

$$= g_{CO} u_D + \frac{g_{CO} \beta}{2} u_D^2 \quad (u_D \leq 0)$$

$$i(t) = g_{CO} V_D \sin(\omega t) - (g_{CO} V_D^2 \alpha / 2) \sin^2(\omega t) \quad (0 \leq t < T'/2)$$

$$= g_{CO} V_D \sin(\omega t) + (g_{CO} V_D^2 \beta / 2) \sin^2(\omega t) \quad (T'/2 \leq t < T')$$

Inserting (A-15) into the definition for the effective value and performing the integrations yields:

$$i_{eff} = \left( g_{CO} V_D \sqrt{2} \right) \left\{ 1 + (3V_D^2 / 32)(\alpha^2 + \beta^2) - (4V_D / 3\pi)(\alpha + \beta) \right\}^{1/2}$$

Inserting (A-16) into (A-9) and remembering that $u_{D_{eff}} = V_D / \sqrt{2}$
we obtain the contribution $g_C^m$ to the measured conductance $g_D^m$ due to
the drain voltage dependent channel conductance:

$$g_C^m = g_{CO} \left\{ 1 + (3V_D^2 / 32)(\alpha^2 + \beta^2) - (4V_D / 3\pi)(\alpha + \beta) \right\}^{1/2}$$

Please note that (A-17) is a function of both drain voltage (through $V_{DO}$) and gate voltage (through $\alpha$ and $\beta$). In Fig. A-8 $g_C^m$ is plotted as a function of $V_{DO}$—the amplitude of the measuring ac-drain voltage—
for a gate voltage $V_G = -0.8V$ which corresponds to the threshold region for the particular IGFET used in this example. For $V_{DO} = 5mV$ for example, $g_C^m$ is $0.96g_{CO}$. Please note that due to the linearization introduced in eq. (A-12) higher harmonics generated by the nonconstant conductance up to second order are considered in (A-17). Of course, higher harmonics are also present but they are neglected in this treatment.

We will now calculate the contribution of the conductance due to the forward biased drain-substrate junction to the measured conductance.

The current flowing between drain and substrate as a function of drain voltage $u_D$ (units of $kT/q$) is given by \(^{(107)}\):

$$i(u_D) = I_s \left[ \exp(u_D) - 1 \right]. \quad (A-18)$$

If $u_D$ is now a sinusoidal ac-drain voltage as given in eq. (A-10) we can write (A-18) as a function of $t$:

$$i(t) = I_s \left[ \exp(u_0 \sin(\omega t)) - 1 \right] \quad (A-19)$$

where $u_0$ is the amplitude of the ac signal in $kT/q$ units. To facilitate the determination of the effective current, we expand $[i(t)]^2$ from (A-19) into a power series which can be written as:

$$[i(t)]^2 = I_s^2 \left[ \exp(2u_0 \sin \omega t) - 2 \exp(u_0 \sin \omega t) + 1 \right] \quad (A-20)$$

$$= I_s^2 \left[ 1 + \frac{2}{1!} u_0 \sin \omega t + \frac{4}{2!} u_0^2 \sin^2 \omega t + \ldots - 2 - \frac{2}{1!} u_0 \sin \omega t \right.$$

$$- \left. \frac{2}{2!} u_0^2 \sin^2 \omega t - \ldots + 1 \right]$$
\[
[i(t)]^2 = I_s^2 \left[ \frac{2}{2!} u_o^2 \sin^2 \omega t + \frac{6}{3!} u_o^3 \sin^3 \omega t + \ldots \ldots + \frac{2^n-2}{n!} u_o^n \sin^n \omega t \right]
\]

(A-20) can also be written as:

\[
[i(t)]^2 = I_s^2 \sum_{n=2}^{\infty} \frac{2^n-2}{n!} u_o^n \sin^n \omega t \quad (A-21)
\]

We insert now (A-21) into the definition of the effective current, eq. (A-11) and obtain:

\[
i_{\text{eff}} = \left\{ \frac{I_s^2}{T'} \int_0^{T'} \sum_{n=2}^{\infty} \frac{2^n-2}{n!} u_o^n \sin^n \omega t \, dt \right\}^{1/2} \quad (A-22)
\]

where \( T' \) is the time period of the ac-signal. The series in (A-22) converges uniformly for all values of \( u_o \) and we can, therefore, interchange summation and integration:

\[
i_{\text{eff}} = \left\{ \frac{I_s^2}{T} \sum_{n=2}^{\infty} \frac{2^n-2}{n!} u_o^n \int_0^{T'} \sin^n \omega t \, dt \right\}^{1/2} \quad (A-23)
\]

We can find now the measured conductance \( g_j^m \) due to the p-n junction conductance \( g_j \) by inserting (A-23) into (A-9), remembering that \( v_{\text{eff}} = kT u_o / (2 q) \) and \( \omega T' = 2\pi \):

\[
g_j^m = \frac{\sqrt{2} q I_s}{kT} \left\{ \frac{1}{2\pi} \int_0^{2\pi} \sum_{n=2}^{\infty} \frac{2^n-2}{n!} u_o^{n-2} \sin^n \theta \, d\theta \right\}^{1/2} \quad (A-24)
\]
Only terms with even \( n \) contribute in the summation of (A-24) because the integrals in (A-24) vanish for odd \( n \). Thus, we can replace \( n \) by \( 2n \) and write:

\[
{g_J^m} = \frac{\sqrt{2} q I_S}{kT} \left\{ \frac{1}{2\pi} \sum_{n=1}^{\infty} \frac{2^{2n-1}}{(2n)!} u_o^{2n-2} \int_0^{2\pi} \sin^{2n} x \, dx \right\}^{\frac{\nu_2}{2}}. \tag{A-25}
\]

The integrals in (A-25) have the following solutions:

\[
\int_0^{2\pi} \sin^{2n} x \, dx = 2\pi \frac{1 \times 3 \times \ldots \times (2n-1)}{2 \times 4 \times \ldots \times (2n)} = 2\pi \prod_{i=1}^{n} \frac{(2i-1)}{(2i)} \tag{A-26}
\]

Inserting (A-26) into (A-25) and canceling common terms of the products in (A-26) and the factorials in (A-25) yields:

\[
{g_J^m} = \frac{\sqrt{2} q I_S}{kT} \left\{ \sum_{n=1}^{\infty} \frac{2^{2n-2} u_o^{2n-2}}{\prod_{i=1}^{n} (2i)^2} \right\}^{\frac{\nu_2}{2}}. \tag{A-27}
\]

It is of advantage to treat the case \( u_o = 0 \) separately. For \( u_o = 0 \) all terms with \( n \neq 1 \) vanish in the series in (A-27) and \( {g_J^m}_{0} \) is given by

\[
{g_J^m}_{0} = \frac{q I_S}{kT} \tag{A-28}
\]

which is identical to (A-7), as expected. The conductance \( {g_J^m} \) relative to the zero drain voltage conductance \( {g_J^m}_0 \) can, therefore, be written as:

\[
{g_J^m} / {g_J^m}_0 = 1 + \sqrt{2} \left\{ \sum_{n=2}^{\infty} \frac{(2^{2n-2}) u_o^{2n-2}}{\prod_{i=1}^{n} (2i)^2} \right\}^{\frac{\nu_2}{2}} \tag{A-29}
\]
where (A-27), (A-28) and the fact that the first term in the series
\( n = 1 \) is independent of \( u_o \) have been used.

In Fig. A-8 equation (A-29) is plotted as a function of \( u_o \).
It can be seen clearly that \( g_J^m/g_J^0 \) increases quite strongly with in-
creasing \( u_o \), especially for \( u_o > 1 \) (\( V_{DO} > kT/q \)).

Fig. A-9 shows a plot of the total measured conductance \( g_D^m \)
as a function of the amplitude of the measuring ac drain voltage with
\( V_G \) as parameter. The curves are plotted relative to the zero drain
voltage conductance \( g_{D0}^m \). \( g_D^m \) considers the influences of both the
drain voltage dependent channel conductance \( g_C^m \) (eq. A-17) and the
p-n junction conductance \( g_J^m \) (eq. A-29) and is given by:

\[
g_D^m/g_{D0}^m = [(g_C^m/g_{CO})g_{CO} + (g_J^m/g_{JO})g_{JO}] / g_{D0}^m \tag{A-30}
\]

where \( g_{D0} = g_{CO} + g_{JO} \) \tag{A-31}

As shown before, \( g_C^m \) and \( g_J^m \) have opposite drain voltage depen-
dence. \( g_C^m \) decreases with increasing ac drain voltage amplitude \( u_o \)
whereas \( g_J^m \) increases with increasing \( u_o \). For gate voltages near thresh-
old \( g_J^m \) dominates the measured channel conductance \( g_D^m \) so that \( g_D^m \) in-
creases with increasing \( u_o \). For larger gate voltages, however, the
influence of \( g_J^m \) on \( g_D^m \) decreases because \( g_C^m \) increases and \( g_J^m \) stays con-
stant. Finally \( g_D^m \) decreases with increasing \( u_o \) when \( g_C^m \) is the domi-
nant term in (A-30). This dependence is clearly shown in Fig. A-9.
Figure A.8: Contribution $g_m^{\text{m}}$ to the measured channel conductance $g_m^{\text{m}}$ due to the drain-substrate p-n junction as a function of drain voltage amplitude $u_o$ (units of $kT/q$).

Figure A.9: Measured channel conductance $g_D^m$ as a function of drain voltage amplitude $u_o$ (units of $kT/q$) and with gate voltage $V_G$ as parameter.
The findings of this chapter show that theoretically the drain conductance should be nearly independent of the measuring ac drain voltage as long as the amplitude of the drain voltage is small or comparable to $kT/q \ 25\text{mV}$ at room temperature. The large dependence of the measured channel conductance as shown in Fig. 5.6 can, therefore, not be explained by the presented theory. The reasons for this failure can be that other mechanisms such as channel length variations, voltage dependent series capacitance, etc., influence the drain characteristics or that the conductance of the drain to substrate junction cannot be described by the Shockley equation (A-7). It is, however, doubtful that these additional effects would be dominant. More reasonable is the assumption that the experimental data are incorrect. Although the findings of Fig. 5.6 were verified on a number of quite different devices, it is still possible that some systematic errors in the measurement techniques existed. It is suggested in the paragraph: "Suggestions for future work" to investigate this interesting effect both experimentally and theoretically in more detail.