1

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

TWO-DIMENSIONAL ANALYTICAL MODELLING OF
THE SHORT-CHANNEL MOSFET

by

James D. Kendall, B. Eng., M. Eng.

A thesis submitted to the Faculty
of Graduate Studies and
Research in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Ottawa-Carleton Institute for Electrical Engineering
Department of Electronics
Faculty of Engineering
Carleton University

October 1987

© copyright
1987, James D. Kendall

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-40614-3

The undersigned recommend to the Faculty of Graduate Studies and Research the acceptance of the thesis:

"Two-Dimensional Analytical Modelling of the Short-Channel MOSFET"

submitted by James D. Kendall in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Prof. A.R. Boothroyd
Thesis Supervisor

Prof. M.A. Copeland
Chairman
Department of Electronics

Prof. C.A.T. Salama
External Examiner
University of Toronto

Carleton University
October 23, 1987

ABSTRACT

The objective of this thesis is to show that accurate models of the short-channel MOSFET can be obtained by two-dimensional analytical solutions of Poisson's equation and the current continuity equation, and that such models are computationally much faster than comparable numerical models. Two models of this type are derived: one general model that is valid in all the normal regimes of operation and another that is specialized for use in only the subthreshold regime. The general model assumes that the substrate is uniformly doped and the mobility is constant, while the subthreshold model allows for an arbitrary substrate doping profile and a field dependent mobility. A two-dimensional Poisson's equation is solved for both models by reducing it to Laplace's equation using a trial function, and solving the resulting Laplace problem by the separation of variables technique. This is sufficient to obtain the subthreshold solution, but for the general case a one-dimensional current continuity equation must be solved simultaneously with Poisson's equation by either an iterative function or Fourier analysis technique. Only the Fourier analysis technique is computationally fast enough to be practical.

Using this technique a closed form expression for the drain current is obtained which is the two-dimensional analogue of the one-dimensional equation derived by Brews [4]. It permits a simple comparison to the numerical simulator MINIMOS [33]

demonstrating that the general analytical model is accurate without the use of fitting parameters, and 5 to 300 times faster than comparable numerical models. The subthreshold solution is used to derive a closed form expression for the short-channel threshold voltage and a subthreshold current model. The former is verified by comparison to MINIMOS simulations and the latter, by comparison to measurements on carefully characterized devices.

The thesis closes with a presentation of partially completed research that shows how the general solution may be extended to include the effects of an arbitrary substrate doping profile and a field dependent mobility. A discussion of using the general solution to develop a short-channel capacitance model, and applying the analytical techniques to model other devices such as the MODFET is also included.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr. Roy Boothroyd, for providing help and encouragement throughout the course of my research. I would also like to thank Dr. Claudine Simson, David Smith and Dr. Cezary Slaby who arranged for Northern Telecom Electronics Ltd. to support my work by supplying devices, characterization facilities, computer time and financial support. Other members of Northern Telecom's staff who deserve thanks are: Alex Kalnitsky for his help with both software and measurement problems, Jaan Kolk for aiding me in the use of his technique for measuring substrate doping profiles, Ian Few for optical measurements of device dimensions, and Gina Mulvie for word processing this manuscript. The contribution of the following faculty members of Carleton University should also be acknowledged: Dr. Garry Tarr for numerous discussions concerning device physics and processing, Dr. Mizanur Rahman for advice on analytical methods for solving partial differential equations, Dr. Lawrence May for help with N-dimensional Newton methods, and Dr. John Goldak for discussions on the finite element method. In addition, I would like to thank the Natural Sciences and Engineering Research Council for their financial support of my work.

Finally, I would like to give special thanks to my wife Catherine Campbell for her aid, moral support and patience throughout the course of this thesis.

TABLE OF CONTENTS

CHAPTER 1	DEFINING THE PROBLEM	1
1.1	Introduction	1
1.2	Literature Review	12
1.3	Objectives of the Thesis	29
1.4	Outline of the Thesis	33
CHAPTER 2	THE SUBTHRESHOLD MODEL	35
2.1	The Poisson Problem and Approximating the Boundary Conditions	35
2.1a	Formulating the Problem	35
2.1b	Boundary Conditions	38
2.1c	Improved Side Boundary Conditions	50
2.1d	Calculating the Cylindrical Depletion Layer Radii	59
2.2	Solution of the Poisson Problem	64
2.3	The Threshold Voltage Model	78
2.4	The Subthreshold Current Model	90
CHAPTER 3	VERIFICATION OF THE SUBTHRESHOLD MODEL	99
3.1	Testing the Threshold Voltage Model	99
3.2	Testing the Subthreshold Current Model	108
3.2a	Determining the Device Parameters	108
3.2b	Measurement of the Substrate Doping Profile	113
3.2c	Triple Agreement	138
CHAPTER 4	SOLUTION OF THE POISSON PROBLEM FOR THE GENERAL MODEL	141
4.1	The Poisson Problem and its Boundary Conditions	141
4.2	The Side Boundary Conditions	145
4.3	The Long-Channel Trial Function	160
4.4	Solution of the Laplace Problem	173
CHAPTER 5	THE ITERATIVE FUNCTION TECHNIQUE	188
5.1	Outline of the Technique	188
5.2	Applying Current Continuity to Calculate the Quasi-Fermi Potential	191
5.2a	Calculating the Quasi-Fermi Potential From the Two-Dimensional Potential	191
5.2b	Calculating the Quasi-Fermi Potential From the Surface Potential	195
5.3	Employing the Continuity of Normal Displacement to Calculate the Surface Potential	201
5.4	The Algorithms	203

5.5	Patterson's Method of Evaluating Fourier Integrals	207
5.5a	Discretizing the Fourier Integrals	207
5.5b	The Chebyshev Coefficients	210
5.5c	The R_{ki}	211
5.5d	Calculating the Spherical Bessel Functions	212
5.5e	A More Efficient Form of Patterson's Method	215
5.6	The Performance of the Iterative Function Technique	217
5.7	Why the Iterative Function Technique Is Too Slow	219
5.8	Summary	226
CHAPTER 6	THE FOURIER ANALYSIS TECHNIQUE	228
6.1	Transformation of the Problem into the Fourier Domain	228
6.2	Newton and Quasi-Newton Methods	233
6.3	Calculation of the Fourier Quantities F_n , G_n and H_n	236
6.4	Details of the Iterative Procedure	244
6.5	Performance of the Fourier Analysis Technique	249
CHAPTER 7	VERIFICATION OF THE GENERAL MODEL	255
CHAPTER 8	PARTIALLY COMPLETED AND FUTURE RESEARCH	265
8.1	Introduction	265
8.2	Improvements to the Subthreshold Model	266
8.2a	The Side Boundary Conditions	266
8.2b	Defining Threshold	269
8.3	Improvements to the General Model	270
8.3a	Including a Field Dependent Mobility	270
8.3b	Including a Nonuniform Substrate Doping Profile	275
8.3c	Other Improvements	285
8.4	Solving for the E Vector	287
8.5	New Areas of Research	289
8.5a	MOSFET Capacitances	289
8.5b	MODFET Modelling	292
CHAPTER 9	CONCLUSION	295
REFERENCES		299

LIST OF FIGURES

<u>FIGURE NUMBER</u>	<u>PAGE</u>
1.1.1 Long and short-channel MOSFETs.	3
1.2.1 The MOSFET geometry with the rectangular domain for the subthreshold solution indicated by cross-hatching.	22
2.1.1 Comparison of the quadratic side boundary conditions to the results of MINIMOS simulations for a uniform substrate device.	43
2.1.2 Comparison of the quadratic side boundary conditions to the results of MINIMOS simulations for an NMOS device.	46
2.1.3 Comparison of the quadratic side boundary conditions to the results of MINIMOS simulations for a CMOS device.	47
2.1.4 Substrate doping profile for the NMOS device.	48
2.1.5 Substrate doping profile for the CMOS device.	49
2.1.6 Limiting the slope at $x = 0$ of the cubic polynomial $C(x)$.	52
2.1.7 Comparison of the cubic side boundary conditions to the results of MINIMOS simulations for a uniform substrate device.	56
2.1.8 Comparison of the cubic side boundary conditions to the results of MINIMOS simulations for an NMOS device.	57
2.1.9 Comparison of the cubic side boundary conditions to the results of MINIMOS simulations for a CMOS device.	58
2.1.10 Depletion layer geometry near the source underdiffusion.	60
2.2.1 Intersection points of $y = \tan u_i$ and the hyperbola $y = d/(R \cdot t_{ox} u_i)$.	71
2.2.2 $T(x)$, $U(x)$, $V(x,y)$ and $\Psi(x,y)$ for a CMOS device where $L = 1 \mu m$.	79

FIGURE NUMBER	PAGE
2.2.3 $T(x)$, $U(x)$, $V(x,y)$ and $\psi(x,y)$ for a CMOS device where $L = 10 \text{ } \mu\text{m}$.	80
3.1.1 Threshold voltage versus channel length for a uniform substrate device.	102
3.1.2 Threshold voltage versus channel length for an NMOS device.	103
3.1.3 Threshold voltage versus channel length for a CMOS device.	104
3.2.1 Circuit used by Kolk's method to measure dV_{SB}/dV_{GS} .	117
3.2.2 Calibration configuration for the circuit in Figure 3.2.1.	118
3.2.3 V_{GS} versus x_d curves for $I_D = I_{D1}$, $I_D = I_{D2}$ and $I_D = 0$.	122
3.2.4 Measured doping profile for Device Type 1.	127
3.2.5 Measured doping profile for Device Type 2.	128
3.2.6 Measured and simulated subthreshold I_D versus V_{GS} characteristics for Device Type 1.	134
3.2.7 Measured and simulated subthreshold I_D versus V_{GS} characteristics for Device Type 2.	135
3.2.8 Drain current versus channel length for measured, analytically simulated and MINIMOS simulated results.	139
4.1.1 The MOSFET geometry with the rectangular domain for the general solution indicated by cross-hatching.	143
4.2.1 Potential profile $\psi(x,L)$ extracted from the results of a MINIMOS simulation.	147
4.2.2 Comparison of the single/double parabola side boundary conditions to the results of MINIMOS simulations for $V_{SB} = 0$.	158
4.2.3 Comparison of the single/double parabola side boundary conditions to the results of MINIMOS simulations for $V_{SB} = 5 \text{ V}$.	159

FIGURE NUMBER	PAGE
4.3.1 Percentage error in the gradual channel approximation and the long-channel surface potential $U(0,y)$ versus position along the channel y for Device 1.	163
4.3.2 Percentage error in the gradual channel approximation and the long-channel surface potential $U(0,y)$ versus position along the channel y for Device 2.	164
5.4.1 Algorithms 1 and 2 for the iterative function technique.	204
5.6.1 Drain current versus iteration number for Device 1 using Algorithm 1 where the mobility is constant in (a) and field dependent in (b).	220
5.6.2 Drain current versus iteration number for Device 1 using Algorithm 2 where the mobility is constant in (a) and field dependent in (b).	221
5.7.1 Three-dimensional plots of $\psi(x,y)$, $U(x,y)$, $V(x,y)$ and $Z(x,y)$ as viewed from the substrate, where with the mobility constant these potentials were solved for using Algorithm 1.	224
5.7.2 Three-dimensional plots of $Z(x,y)$ as viewed from the silicon surface.	225
6.4.1 Flowchart of the Fourier analysis technique when the mobility is constant.	247
6.5.1 Drain current versus iteration number for Device 1 biased below threshold.	250
6.5.2 Drain current versus iteration number for Device 1 biased just above threshold.	251
6.5.3 Drain current versus iteration number for Device 1 biased well above threshold.	252
7.1 I_D versus V_{DS} for Device 1 where $V_{SB} = 0$.	258
7.2 I_D versus V_{DS} for Device 2 where $V_{SB} = 0$.	259
7.3 I_D versus V_{DS} for Device 1 where $V_{SB} = 2$ V.	260
7.4 I_D versus V_{GS} for Device 1.	262

FIGURE NUMBERPAGE

- 7.5 I_D versus V_{GS} for Device 2. 263
- 8.2.1 (a) Comparison of the cubic drain side boundary condition and (b) the single/double drain side boundary condition to the results of MINIMOS simulations for a device whose parameters are $L = 3 \text{ } \mu\text{m}$, $W = 24 \text{ } \mu\text{m}$, $N_B = 5 \times 10^{15} \text{ cm}^{-3}$, $N_S = 2.3 \times 10^{20} \text{ cm}^{-3}$, $t_{ox} = 500 \text{ \AA}$, $r_j = 0.5 \text{ } \mu\text{m}$, $V_{FB} = -0.50 \text{ V}$ and whose applied voltages are $V_{GS} = 0.5 \text{ V}$, $V_{DS} = 5 \text{ V}$, $V_{SB} = 0$. 267.
- 8.2.2 (a) Comparison of the cubic drain side boundary condition and (b) the single/double drain side boundary condition to the results of MINIMOS simulations for a device whose parameters are $L = 1 \text{ } \mu\text{m}$, $W = 24 \text{ } \mu\text{m}$, $N_B = 2 \times 10^{16} \text{ cm}^{-3}$, $N_S = 2.3 \times 10^{20} \text{ cm}^{-3}$, $t_{ox} = 500 \text{ \AA}$, $r_j = 0.532 \text{ } \mu\text{m}$, $V_{FB} = -0.72 \text{ V}$ and whose applied voltages are $V_{GS} = 0.5 \text{ V}$, $V_{DS} = 5 \text{ V}$, $V_{SB} = 0$. 268
- 8.3.1 Flowchart of the Fourier analysis technique when the mobility is field dependent. 274
- 8.5.1 Physical structure of the MODFET reproduced from [71]. 293

LIST OF TABLES

<u>TABLE NUMBER</u>		<u>PAGE</u>
2.1.1	Device parameters for the uniform substrate, NMOS and CMOS devices.	44
3.2.1	Measured values of N_B , $D_I x_C$, D_I and x_C for Device Types 1 and 2.	129
3.2.2	Ratios of drain current and channel length versus channel length with fixed charge values obtained from fitting measured and simulated subthreshold currents for Device Type 1.	136
3.2.3	Ratios of drain current and channel length versus channel length with fixed charge values obtained from fitting measured and simulated subthreshold currents for Device Type 2.	136
4.1.1	The matching depth x_m as a function of the substrate doping N_B and junction depth r_j as extracted from MINIMOS simulations.	155
4.3.1	Parameters for Devices 1 and 2.	162
7.1	Parameters for Devices 1 and 2. (This table is identical to Table 4.3.1 and is repeated here for convenience.)	256

LIST OF SYMBOLS

a	$= d/(R t_{ox})$ a parameter used in the solution of the eigenvalues u_i
a	a variable used in the solution of U_s
A	a constant used in the derivation of Gummel's algorithm.
A	$= N_B/(2C)$ a parameter used in the subthreshold initial guess for the surface potential
a_C	a variable used in the solution of U_s
a_i	a Chebyshev coefficient
A_i	a coefficient in the Fourier series for $V(x,y)$
A_m	a variable used to evaluate continued fractions
A_n	a Fourier coefficient used to evaluate H_n
a_0, a_1, a_2	coefficients used in the single/double parabola side boundary conditions
$A(S_s)$	the antiderivative function of $Q_m(S'_s)$
b	a variable used in the solution of U_s
b_m	a variable used to evaluate continued fractions
B_i	a coefficient in the Fourier series for $V(x,y)$
B_m	a variable used to evaluate continued fractions
B_n	a Fourier coefficient used to evaluate H_n
b_0, b_1, b_2	coefficients used in the single/double parabola side boundary conditions
$b_{00}-b_{22}$	coefficients needed by the interpolation polynomial for x_m
c	a variable used in the solution of U_s
c_s	a variable used in the theorem that states how Fourier sine series with nonzero endpoints are to be differentiated

C	$= C_{ox}^2 / (2q\epsilon_s)$ a parameter used to solve for the surface potential
C_{gb}	the measured gate to bulk capacitance
C_i	a coefficient used in the threshold voltage model
C_n	a Fourier coefficient used to solve for H_n
C_{ox}	the capacitance per unit area of the gate oxide
C_{pad}	the capacitance of a probe pad
C_D	capacitance per unit area of the depletion region
C_K	a constant used to calculate the mobile charge density
C_0, C_1, C_2	coefficients used in the single/double parabola side boundary conditions
$C(x)$	a cubic polynomial used in the cubic side boundary conditions
d	the depth of the rectangular region within which Poisson's equation is solved
d	<u>a variable used in the solution of U_s</u>
d_s	a variable used in the theorem that states how Fourier sine series with nonzero endpoints are to be differentiated
D	a variable used to calculate $d^2 U_s / dy^2$
D_i	a coefficient used in the threshold voltage model
D_n	a Fourier coefficient used to solve for H_n
D_I	the dose of the nonuniformity in the doping profile
d_0, d_1, d_2	coefficients used in the single/double parabola side boundary conditions
$\underline{E}$	the vector of the Fourier coefficients E_n
E_{cx}	a parameter that specifies the dependence of mobility on the x component of electric field

E_{cy}	a parameter that specifies the dependence of mobility on the y component of electric field
E_g	the bandgap
E_i	a coefficient used in the threshold voltage model
E_n	a coefficient in the Fourier series for $Z(x,y)$
E_s	the electric field at the silicon surface
E_x	the x component of electric field
E_y	the y component of electric field
f	a function used in Gummel's algorithm
F_i	a coefficient used in the threshold voltage model
F_n	a Fourier coefficient used to solve for H_n
F_D	the correction to the normal electric field at the point (0,L) so that the electric displacement is made continuous for the cubic side boundary condition
F_S	the correction to the normal electric field at the point (0,0) so that the electric displacement is made continuous for the cubic side boundary condition
$f(x)$	a function related to the nonuniform part of the doping profile
$f(\mathbf{r})$	a function used to calculate Q_m
$f_n(\underline{E})$	a function of the vector $\underline{E}$ used in the N-dimensional Newton's method solution for $\underline{E}$
$f_D(x)$	a function specifying the single/double parabola boundary condition for the drain side
$f_S(x)$	a function specifying the single/double parabola boundary condition for the source side
$F(x)$	the indefinite integral of $f(x)$
G_n	a Fourier coefficient used to solve for H_n
G_D	a variable used in the threshold voltage model related to F_D

G_S	a variable used in the threshold voltage model related to F_S
$g(x,y)$	a function which is the logarithm of the sum of two exponentials
$g(y)$	a function used with the Fourier analysis technique
$g(y_m)$	a function used in the calculation of y_m
$G(x)$	the second indefinite integral of $f(x)$
H	the inverse of the Jacobian matrix
H_n	a Fourier coefficient needed to calculate the E_n
H_0	the fundamental Fourier coefficient of the H_n from which the drain current can be calculated
$h(d)$	a function used to calculate the depletion depth d
$h(y)$	a function used with the Fourier analysis technique
i	an integer
I_{cn}	a symbol representing the Fourier cosine transform integral
I_D	the drain current
I_{D1}	the smaller drain current used in Kolk's method
I_{D2}	the larger drain current used in Kolk's method
I_{sn}	a symbol representing the Fourier sine transform integral
j	an integer
J	the Jacobian matrix
$j_k(w)$	the k 'th spherical Bessel function
k	an integer
k	Boltzmann's constant
K	$= qN_B/\epsilon_S$ a parameter that depends upon the substrate doping
K	a constant used in the derivation of Kolk's method

L	the electrical channel length
L_{min}	the minimum channel length that can still be considered long-channel
L_D	the Debye length
L'	the effective reduced channel length in the saturation regime
m	an integer
M	$= qN_B / (2\epsilon_s)$ a parameter that depends upon the substrate doping
n	an integer
n_i	the density of carriers in an intrinsic semiconductor
N	the number of Fourier coefficients E_n
N_A	the density of acceptor atoms
N_B	the constant doping density in the substrate
N_{Be}	the effective constant doping density in the substrate used to calculate R_S and R_D when the substrate is nonuniformly doped
N_C	the number less one of the Chebyshev coefficients, polynomials or collocation points used in a Chebyshev expansion
N_D	the density of donor atoms
N_S	the highest doping density in the source and drain regions
N_1, N_2	doping densities used to express the nonuniform doping profile determined using Kolk's method
$N(x)$	the substrate doping profile
$N_n(x)$	the nonuniform part of the substrate doping profile
p	a parameter used to determine the trial function $U(x)$
P	a constant used to determine U_s

P_n	a Fourier coefficient used by the Fourier analysis technique if the mobility is field dependent
$p(L')$	a function used to determine L'
$P_k(u)$	the k 'th Legendre polynomial
q	the charge on an electron
Q	a constant used to determine U_s
Q_f	the density of fixed charge per unit area at the silicon-silicon dioxide interface
Q_m	the density of mobile charge per unit area in the inversion layer
Q_{mLC}	the part of Q_m that would be found in an ideal long-channel device
Q_{mSC}	the part of Q_m that is induced due to short-channel effects
Q_B	the total charge in the depletion region
Q_D	the density of charge per unit area in the depletion region
Q_G	in the derivation of Kolk's method this is the density of charge per unit area on the gate while in the explanation of short-channel capacitances it is total charge on the gate
Q_M	the total mobile charge in the inversion layer
r	the radial distance in a polar coordinate system
r_j	the junction depth
R	$= \epsilon_s / \epsilon_{ox}$ the ratio of the dielectric constants of silicon and silicon dioxide
R_k	the ratio of two consecutive spherical Bessel functions evaluated at the same argument
R_{ki}	the integral of the product of the k 'th Legendre polynomial $P_k(u)$ and the i 'th Chebyshev polynomial $T_i(u)$ over the interval $[-1,1]$
R_q	a variable used to determine U_s

R_B	a constant used to determine U_S
R_D	the radius of the cylindrical depletion edge near the drain underdiffusion
R_S	the radius of the cylindrical depletion edge near the source underdiffusion
$R(x,y)$	the sum of $V(x,y)$ and $Z(x,y)$
$\underline{s}$	the difference vector between E vectors on two consecutive iterations of the Fourier analysis technique used by Broyden's method to estimate the H matrix
S_S	the long-channel surface potential, i.e. $S(0,y)$
S_{SD}	the long-channel surface potential at point $(0,L)$
S_{SS}	the long-channel surface potential at point $(0,0)$
S_{Cin}	the matrix of sums used for the high speed numerical evaluation of the Fourier cosine integral
S_{Sin}	the matrix of sums used for the high speed numerical evaluation of the Fourier sine integral
$S(x,y)$	the sum of the trial functions $T(x,y)$ and $U(x,y)$ which is the two-dimensional potential in an ideal long-channel device
t	a dummy integration variable
t_{ox}	the gate oxide thickness
T	the absolute temperatures in degrees Kelvin
T	a variable used to calculate Q_m
T_D	the x derivative of potential at point $(0,L)$
T_S	the x derivative of potential at point $(0,0)$
$T(x)$	the one-dimensional nonuniformity trial function
$T(x,y)$	the two-dimensional nonuniformity trial function
$T_i(u)$	the i 'th Chebyshev polynomial
u	a variable used in the evaluation of $g(x,y)$

u	a substitution variable used for transforming integrals
u_i	the eigenvalues required by the Fourier series in $V(x,y)$
u_j	the Chebyshev collocation points
u_0	the fundamental eigenvalues of the u_i
U_s	the long-channel surface potential in a uniform substrate device, i.e. $U(0,y)$
U_{SD}	the potential $U(0,L)$
U_{SS}	the potential $U(0,0)$
U_x	the x derivative of $U(x,y)$ along the silicon surface, i.e. $\partial U / \partial x (0,y)$
$U(x)$	the one-dimensional uniformity trial function
$U(x,y)$	the two-dimensional uniformity trial function
V_s	the potential $V(0,y)$
V_x	the x derivative of $V(x,y)$ along the silicon surface, i.e. $\partial V / \partial x (0,y)$
V_X	the RMS voltage obtained from the measurement circuit in Kolk's method
V_C	a constant equal to 0.15 V which is the assumed voltage drop between the source contact and the point $(0,0)$, and the drain contact and the point $(0,L)$
V_D	the potential at the drain contact
V_{DS}	the applied drain to source voltage
V_{FB}	the flatband voltage
V_{GB}	the applied gate to substrate voltage
V_{GB}	the effective gate voltage as modified by the flatband voltage, i.e. $V_{GB} - V_{FB}$
V_{GBCr}	the critical gate voltage below which the mobile charge is negligible

V_{GBCrD}	the critical gate voltage at point (0,L)
V_{GBCrS}	the critical gate voltage at point (0,0)
V_{GBT}	the threshold voltage referenced to the substrate which is used to calculate the surface potential
V_{GS}	the applied gate to source voltage
V_L	the voltage V_{MD} less $T(0)$
V_{MD}	the approximate potential at point (0,L) assumed to be $V_D - V_C$
V_{MS}	the approximate potential at point (0,0) assumed to be $V_S - V_C$
V_S	the potential at the source contact
V_{SB}	the applied source to substrate voltage
V_{SBmax}	the maximum value of V_{SB} that can be applied without breaking down the source-substrate junction
V_{SC}	a voltage function related to Q_{MSC} by a constant
V_T	the threshold voltage
V_0	the voltage V_{MS} less $T(0)$
$V(x,y)$	the solution of the Laplace problem that accounts for the fact that the source and drain raise the potential of the substrate in their vicinity
$V(y)$	the difference between the quasi-Fermi-potentials
W	the width of the channel
W_D	the planar depletion width of the drain junction
W_S	the planar depletion width of the source junction
x	x coordinate in the Cartesian coordinate system
x_d	the depletion depth
x_m	the depth at which the two parabolas are matched in the double parabola boundary condition

x_0	a generic variable that can represent either x_S or x_D
x_C	the centroid of the nonuniformity in the doping profile
x_D	the distance from the point $(0, L)$ to the point on the depletion edge that is directly below it
x_S	the distance from the point $(0, 0)$ to the point on the depletion edge that is directly below it
$X_i(x)$	function used in the separation of variables solution for $V(x, y)$
$X_n(x)$	function used in the separation of variables solution for $Z(x, y)$
y	y coordinate in the Cartesian coordinate system
Δ	the difference vector between $\underline{f}(E)$ vectors on two consecutive iterations of the Fourier analysis technique used by Broyden's method to estimate the H matrix
Y_m	the y coordinate of the minimum surface potential
$Y_i(y)$	function used in the separation of variables solution for $V(x, y)$
$Y_n(y)$	function used in the separation of variables solution for $Z(x, y)$
z	a substitutional variable used in the derivation of Gummel's algorithm
z	a substitutional variable used in the derivation of a practical method to calculate $g(x, y)$
Z_S	the potential $Z(0, y)$
Z_x	the x derivative of $Z(x, y)$ along the silicon surface, i.e. $\partial Z / \partial x (0, y)$
$Z(x, y)$	the solution of the Laplace problem that accounts for the effect of mobile charge induced by short-channel effects
α	a parameter used to describe the dependence of mobility on the y component of electric field

α_i	a variable used to specify $\nabla(x,y)$
β	a constant equal to $q/(kT)$
γ	a variable used to damp the N-dimensional Newton's method
$\delta_{ik}, \delta_{nk}, \delta_{i,(k+n)}, \delta_{i,(k-n)}, \delta_{D1}, \delta_{D2}, \delta_{S1}, \delta_{S2}$	Kronecker deltas that have a value of either 1 or 0 depending upon some condition
ΔL	the channel shortening in saturation
ΔV_{FB}	adjustment to the flatband voltage necessary to correct the discrepancy in equation (3.2.2)
ΔV_T	the short-channel threshold voltage reduction
ϵ_{ox}	the dielectric constant of silicon dioxide
ϵ_s	the dielectric constant of silicon
θ	a dummy integration variable
$\lambda(y), \lambda_c(y)$	functions used by Gummel's algorithm to calculate $\phi_n(y)$
μ	the carrier mobility
μ^*	the effective value of carrier mobility
μ_n	the electron mobility
μ_{no}	the bulk, low field value of electron mobility
μ_n^*	the effective value of electron mobility
π	the constant 3.141592654 ...
ρ	the charge density per unit volume
σ_1, σ_2	standard deviations used by Gaussian functions to express the nonuniform doping profile determined using Kolk's method
$\Sigma_1 - \Sigma_9$	sums of trigonometric series that are used by the threshold voltage and subthreshold current models
ϕ	the sum of the quasi-Fermi potentials

ϕ_n	the electron quasi-Fermi potential
ϕ_{nD}	the electron quasi-Fermi potential at the drain
ϕ_p	the hole quasi-Fermi potential
ϕ_{po}	the hole quasi-Fermi potential at the silicon surface
ψ	the electrostatic potential
ψ_{bi}	the built-in potential across a pn junction
ψ_{Cr}	the critical surface potential below which electron charge is negligible
ψ_{CrD}	the critical surface potential at the point (0,L)
ψ_{CrS}	the critical surface potential at the point (0,0)
ψ_m	the potential at the point where the two parabolas are matched in the double parabola boundary condition
ψ'_m	the derivative of the potential with respect to x at the point where the two parabolas are matched in the double parabola boundary condition
ψ_s	the surface potential $\psi(0,y)$
ψ_{sD}	the potential $\psi(0,L)$
ψ_{sL}	the potential $\psi(0,L)$
ψ_{sS}	the potential $\psi(0,0)$
ψ_{s0}	the potential $\psi(0,0)$
ψ_x	the x derivative of $\psi(x,y)$ along the silicon surface, i.e. $\partial\psi/\partial x (0,y)$
ψ_{xL}	the derivative $\partial\psi/\partial x (0,L)$
ψ_{x0}	the derivative $\partial\psi/\partial x (0,0)$
w	the argument of the spherical Bessel functions

TWO-DIMENSIONAL ANALYTICAL MODELLING OF THE SHORT-CHANNEL MOSFET

CHAPTER 1

DEFINING THE PROBLEM

1.1 Introduction

As the physical dimensions of a MOSFET are reduced, several important advantages are realized. Two of these advantages are that the current sourcing capability of the device is increased, and its gate capacitance is reduced. Both effects serve to increase circuit speed. Another advantage is that when the size of the MOSFET is reduced it is possible to have more devices per unit area, which allows higher levels of integration. Thus reducing the dimensions of the MOSFET permits the fabrication of faster, more complex integrated circuits. One of the major obstacles that stands in the way of this goal is that a physically based, analytical model of the electrical behavior of the small geometry MOSFET does not exist. This is due to the fact that the one-dimensional analytical models developed for the large geometry MOSFET [1-4] fail to predict the correct behavior of the small geometry MOSFET because of its fundamentally two-dimensional or even three-dimensional nature.

The small geometry effects of the MOSFET fall into two categories: short-channel effects that are due to reducing the distance between the source and drain; and narrow width effects that are due to reducing the width of the gate. They are similar

in that each represents the influence of edge effects on the behavior of the device. But since it would be too complicated to consider both at once, short-channel effects are chosen as the subject of this thesis and narrow-width effects will not be considered further.

The physical differences between a long-channel and a short-channel MOSFET are illustrated in Figure 1.1.1. For general reference a device may be considered "short-channel" if its channel length is of comparable magnitude to its drain and source junction depths, depletion layer widths and oxide thickness. A device is considered "long-channel" if its channel length is much larger than these quantities. A more precise definition of the division between long and short-channel MOSFETs is expressed by the following equation that was extracted from measured and numerically simulated data by Brews et al. [5].

$$L_{\min} = 0.4 [r_j t_{\text{ox}} (W_S + W_D)^2]^{1/3} \quad (1.1.1)$$

where r_j is the source and drain junction depth, t_{ox} is the oxide thickness, and W_S and W_D are the planar depletion layer thicknesses of the source and drain respectively. The quantity $L_{\min}$ is the minimum channel length that can still be considered long-channel. Hence any device with a channel length shorter than $L_{\min}$ is deemed to be short-channel. It is important to note that equation (1.1.1) defines whether a device is electrically short-channel, which can be quite different from whether it is physically short-channel. For example, a particular device with a $0.5 \mu\text{m}$ channel length could be considered long-channel, and

Long-Channel MOSFET

Short-Channel MOSFET

Figure 1.1.1. Long and short-channel MOSFETs.

another device with a 5,0 μm channel length could be considered short-channel, depending on how the parameters of these devices are chosen.

The difference in the analysis of the electrical behavior of short and long-channel MOSFETs lies in the approximations that may be made. If no approximations are made then the complete solution consists of simultaneously solving a two-dimensional Poisson's equation and a two-dimensional current continuity equation. The results of this solution are the two-dimensional electrostatic potential and the two-dimensional hole and electron quasi-Fermi potentials. From these can be obtained the currents, conductances, charge distributions, stored charges and capacitances of the device.

An approximation that can be made to simplify this problem is to restrict the current flow to one dimension and hence a one-dimensional current continuity equation need only be solved. This is a reasonable approximation as long as the carrier transport is exclusively, or at least mainly, along the silicon surface. Such an approximation can even be used in saturation because although the carriers leave the surface near the drain in this regime, and hence have two-dimensional transport, their concentration in this area is very small and they have little effect on the electrostatic potential. Thus the carriers can be artificially restrained to flow in one dimension along the surface with little net effect on the final solution. It has been found [6-9] that in the normal regimes of operation

(i.e. subthreshold, triode and saturation) an accurate representation of the drain current and capacitance characteristics of the short-channel MOSFET can be obtained using this approximation. Hence for short-channel modelling it is sufficient to solve a two-dimensional Poisson's equation and a one-dimensional current continuity equation.

The problem can be simplified even more by reducing the two-dimensional Poisson's equation to one dimension by employing the gradual channel approximation. This approximation requires that:

$$\frac{\partial^2 \psi}{\partial y^2} \ll \frac{\partial^2 \psi}{\partial x^2} \quad (1.1.2)$$

where $\psi(x,y)$ is the electrostatic potential and the coordinate system is as defined in Figure 1.1.1. Thus Poisson's equation is reduced to its one-dimensional form.

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} \approx \frac{\partial^2 \psi}{\partial x^2} = \frac{-\rho(x,y)}{\epsilon_s} \quad (1.1.3)$$

where $\rho(x,y)$ is the charge distribution and ϵ_s is the permittivity of silicon. The gradual channel approximation is valid for the long-channel MOSFET throughout the channel except near the source and drain boundaries. If the channel length is sufficiently great, then it is reasonable to replace these boundary regions by idealized boundaries where the gradual channel approximation is taken to be valid. In other words, the

two-dimensional regions near the source and drain are regarded as being negligibly small. It is then possible to obtain a model of the "ideal long-channel MOSFET" where the gradual channel approximation is assumed to apply throughout the channel, by simultaneously solving a one-dimensional Poisson's equation and a one-dimensional current continuity equation [1-4]. Once the problem is reduced to one dimension in this fashion it is simple enough to be solved by analytical means.

In contrast, it is not possible to assume that the gradual channel approximation applies throughout the channel for a short-channel MOSFET. This is because the regions where it does not apply near the source and drain are of non-negligible extent, as they occupy most or all of the channel. Therefore a two-dimensional Poisson's equation must be solved to determine the behavior of a short-channel MOSFET, not a one-dimensional Poisson's equation as for a long-channel MOSFET. This is the fundamental difference between long and short-channel modelling. As mentioned previously, the same sort of restriction does not apply to the current continuity equation as it is permissible to use the one-dimensional version of this equation for the short-channel MOSFET provided that the current flows primarily along the surface.

It is however possible to modify the ideal long-channel MOSFET model so that it will give a reasonably accurate representation of the electrical properties of the short-channel MOSFET. This is done by using an approximate short-channel threshold voltage model like those of Yau [10], Taylor [11], and

Ratnakumar and Meindl [12], and a channel shortening model like those of Reddi and Sah [21], Fröhman-Bentchkowsky and Grove [22], and Elmansy and Boothroyd [23]. The short-channel threshold voltage reduction ΔV_T given by the threshold voltage models is used to calculate an effective value of substrate doping [13-15], an effective gate voltage [16], or an effective substrate bias [17]. The rationale for this is that in a short-channel MOSFET the electric displacement from the source and drain regions terminate on a significant amount of depletion charge underneath the gate, thereby reducing the quantity of depletion charge in the substrate "seen" by the gate. This effect is called "charge sharing" [10,11] and can be modelled by using an effective substrate doping, gate voltage or substrate bias as a modification to a long-channel model. The channel shortening ΔL calculated using the second set of models, is the length of the region near the drain where the gradual channel approximation does not apply. There is a similar region near the source but it is very small and is usually ignored because it is not critical to calculating the electrical characteristics. If the channel is shortened by ΔL at the drain end, and the source end is approximated as a region that obeys the gradual channel approximation, then what results is an ideal long-channel device with a reduced channel length. The ideal long-channel model can then be applied to calculate the electrical characteristics of the reduced channel length device, but the characteristics will be modified by the bias voltage dependencies of the channel

shortening ΔL . In particular, a finite conductance in the saturation region is provided by ΔL .

Since the technique of modifying the long-channel model so that it can simulate short-channel devices is approximate, fitting parameters must be introduced into the model to fit the simulated characteristics to measured characteristics. This assumes that the devices under consideration have been fabricated and are available for measurement which may not be the case. Thus the ideal long-channel models that have been modified to handle short-channel effects are limited in the scope of their application by their dependence on fitting parameters. To develop a short-channel MOSFET model which can be used in applications where fitting parameters are not available, it is necessary to dispense with the gradual channel approximation and solve the two-dimensional Poisson's equation. Furthermore, to produce such a model that has minimal computational requirements, analytical methods must be used to solve Poisson's equation.

Before further examination and classification of MOSFET models, it is useful to study the intended applications for these models. These applications are: process design, device design, circuit design, statistical modelling, and simulation for the purpose of basic physical understanding of device operation. Process design establishes the steps that will be followed to fabricate the MOSFETs, their interconnections and any other devices on the integrated circuit. A device model is used to check the electrical behavior of the device that results for a

given process. Device design is the selection of the device parameters and structure in order to give optimum electrical performance. Clearly this requires a device model to predict the electrical characteristics from the device structure. Circuit design involves the simulation of a circuit's behavior using a circuit simulator that relies upon the device model to specify the electrical behavior of the device for a given set of biases. Circuit simulation is usually done for devices whose electrical characteristics are well known because they have been fabricated by an established process. Hence measured data is available from which to extract fitting parameters. But this is not always so as circuit simulation can be used as an aid in process and device design and for this case fabricated devices are not available. Statistical modelling is either done to discover how a variation in process parameters (e.g. furnace temperature and time, implant dose and energy, etc.) affect device and circuit performance, as is done by the FABRICS program [18-19], or to determine the variation in process parameters from device and circuit measurements, as is done by the CASTAM program [20]. In either case a device model is essential to predict the electrical characteristics from the device structure resulting from the processing steps and this device model must be valid over a wide range of device parameters. Lastly, simulation for the purpose of basic physical understanding of device operation is done using a detailed device model that provides information about potentials, fields, charge distributions and current densities. Information gained from this sort of simulation allows one to

make intelligent choices in process and device design and understand the physical causes behind unexpected device behavior. All of these device model applications, except for circuit simulation for an existing process, involve predicting electrical behavior from the device structure without the use of fitting parameters. Thus for these applications a complete solution of a two-dimensional Poisson's equation and a one or two-dimensional current continuity equation is necessary if short-channel MOSFETs are to be modelled, while for circuit simulation a modified long-channel model employing fitting parameters will suffice.

The difficulty with using a model that solves a two-dimensional Poisson's equation and a one or two-dimensional current continuity equation, is that the solution must be done using the computation intensive numerical techniques of finite elements or finite differences. Both techniques involve dividing the physical domain under consideration into small elements, that are usually triangular or rectangular, and then finding the values of the solution function (e.g. the electrostatic potential) only at the nodes of these elements. With finite differences this is done by converting the partial differential equations under consideration, which for a MOSFET are Poisson's equation and the current continuity equation, into difference equations, and then solving these by numerical techniques. The finite element method first expands the solution function using Lagrangian interpolation functions defined on each element, substitutes this expansion into the partial differential equations, and then solves for the coefficients of the Lagrangian

expansion. Using either technique the resulting short-channel MOSFET model requires between several seconds and several days of mainframe CPU time per bias point, depending upon the complexity of the problem and the efficiency of the code. By comparison modified long-channel MOSFET models that have been designed for high speed circuit simulators generally require several tenths of a millisecond of CPU time per bias point. Despite their high computation costs, numerical simulators have a number of advantages and have proved to be very useful in certain applications. Since they can handle arbitrary geometries, boundary conditions, and device structures, they are very useful for investigating the basic physics of device operation and doing device design. But for doing even simple circuit simulation or statistical modelling, where solutions need to be obtained at a large number of bias points, numerical models have prohibitive computing costs.

It is clear that all five categories of device model applications would benefit from a model that would solve the two-dimensional Poisson's equation and one or two-dimensional current continuity equation without incurring the large computational costs of finite elements or finite differences. This would permit two-dimensional device models to be used in circuit simulation and statistical modelling, as well as increase the speed and ease of doing process and device design. The method by which such a model can be created is to apply sophisticated two-dimensional analytical techniques to solve Poisson's equation and the current continuity equation, thereby

creating a new class of device models. Such is the approach of this thesis. Models derived in this manner are similar to the one-dimensional analytical models in that they use analytical techniques together with certain approximations related to the boundary conditions. They are also similar to the numerical models in that they can have a unified electrical and physical analysis and do not require fitting parameters. Thus the two-dimensional analytical models represent a compromise between one-dimensional analytical models and numerical models, and consistent with this they have intermediate computing requirements. The development of two-dimensional analytical models has been impeded by the sheer complexity of the analytical mathematics that is needed. But in this mathematical complexity lies the reason for the potential speed capabilities of such models. Namely, the more of the solution process that is done analytically, the less the numerical demand on the computer. The exact nature of the two-dimensional analytical models that will be developed in this thesis will be made clear after the historical perspective that is added by the literature review is examined.

1(2) Literature Review

The evolution of MOSFET models to deal with devices of ever smaller channel lengths has been by: (1) increasingly sophisticated modifications of one-dimensional analytical ideal long-channel theory; by (2) numerical solution of the semiconductor equations in two dimensions; or by (3) analytical

solutions in two dimensions. In general, in the case of (1), a complete modified long-channel model capable of simulating short-channel MOSFET behavior is comprised of three parts: a drain current expression for an ideal long-channel MOSFET where both channel shortening and charge sharing are neglected; a model to calculate the channel shortening ΔL ; and a short-channel threshold voltage model to calculate the reduction in the threshold voltage ΔV_T due to charge sharing. Several different models are available for each of these effects and to make a complete MOSFET model one simply chooses a model from each of the three categories.

The first expression for the drain current of an ideal long-channel MOSFET was due to Ihantola and Moll [1]. It results from a simple, heuristic analysis that models the surface potential as $2\phi_p + V(y)$ where ϕ_p is the hole quasi-Fermi potential in the neutral bulk (assuming the device is n-channel) and $V(y)$ is the separation of the quasi-Fermi potentials along the silicon surface. This model makes use of the sheet charge approximation which represents the inversion layer as being of zero thickness. The surface potential approximation underestimates the inversion charge but eliminates the need to calculate the boundary surface potential at each end of the channel. This model was superseded by that of Pao and Sah [2] which undertakes a much more sophisticated analysis of the inversion layer. It uses the complete one-dimensional solution of Poisson's equation that includes the effect of depletion, hole and electron charges [24], and an iterative calculation of the

surface potential in a double numerical integral to calculate the drain current. It also assumes a constant mobility. This model is very accurate for long-channel MOSFETs, but requires a large amount of computation due to its reliance on numerical techniques. It was later shown by Pierret and Shields [3] that the amount of computation can be greatly decreased by reducing the double numerical integral to two single numerical integrals. A simpler, but still fairly accurate long-channel drain current expression was derived by Brews [4], who, following Ihantola and Moll, approximated the inversion layer as a sheet of charge, and found the inversion charge by subtracting the depletion charge from the gate displacement. This produces an analytically integrable solution whose only numerical requirements are two iterative solutions for the potentials at the ends of the channel.

An expression to calculate the channel shortening in saturation ΔL was first derived by Reddi and Sah [21] who solved a one-dimensional Poisson's equation between the pinch-off point and the drain, set the component of electric field parallel to the surface at the pinch-off point to zero, and thereby calculated the channel shortening. This model is not very accurate but with a nonzero estimate of the field at the pinch-off point it gives reasonable results [13-15]. A more accurate model was developed by Frohman-Bentchkowsky and Grove [22], who solved three one-dimensional Poisson's equations for the gate to channel, channel to drain, and drain to gate interactions. The resulting three electric field components are

combined using fitting parameters to find the electric field at the pinch-off point, which is then used to calculate the channel shortening. What is unsatisfactory about these two channel shortening models is that they use one-dimensional Poisson's equation solutions to model a phenomenon which is really two-dimensional. El-Mansy and Boothroyd [23] succeeded in calculating the channel shortening from a two-dimensional analysis by employing the integral equivalent of Poisson's equation known as Gauss's Law, which equates the surface integral of normal electric displacement to the total charge contained within the volume. They calculated the channel shortening by choosing a rectangular region between the pinch-off point and the drain, calculating an approximate field pattern around the rectangle, then choosing the channel shortening so that it balanced displacement and charge.

The short-channel threshold voltage reduction ΔV_T is usually modelled in modified long-channel models using the results of Yau [10], Taylor [11] or Ratnakumar and Meindl [12]. The Yau and Taylor threshold voltage models are charge sharing models that use a trapezoid to partition the depletion charge beneath the gate in an arbitrary fashion. The charge partition is used to find the effective substrate doping seen by the gate, and from this the short-channel threshold voltage is calculated. They are essentially the same model except that Yau neglected drain voltage dependence which was added by Taylor. The Ratnakumar and Meindl threshold voltage model is based upon a two-dimensional analytical solution of Poisson's equation, and so has the

potential to produce a physically accurate model. But it uses very approximate boundary conditions in the solution and the Fourier series from which the threshold voltage expression results is truncated to one term. This produces a threshold voltage expression that has the correct functional dependencies but which must employ fitting parameters to be accurate. This model will be discussed in more detail later in the literature review.

A modified long-channel MOSFET model that will give a reasonably accurate representation of the behavior of a short-channel MOSFET with the use of fitting parameters, can be constructed by combining one of each of the drain current, channel shortening and short-channel threshold voltage models. For example the Boothroyd model [16] employs the Brews sheet charge expression for the drain current, calculates the channel shortening from the El-Mansy and Boothroyd model, and adds the threshold reduction ΔV_T to the gate voltage based upon a combined Yau-Ratnakumar and Meindl threshold voltage model. Accounting for the charge sharing effect by adding ΔV_T to the gate voltage is an unusual but useful step. The more common practice of calculating an effective substrate doping [13-15] introduces two substrate doping levels into the model: the real one and the effective one; which is confusing. A popular set of choices, probably due to their analytical simplicity, is to use the Ihantola and Moll expression for the drain current, a modified Reddi and Sah model for the channel shortening, and a Yau model for the threshold voltage reduction to calculate an effective

substrate doping. This is the approach taken by Chatterjee and Leiss [13], Dng [15], and Silburt, Foss and Petrie [17]. Guebels and Van De Wiele [14] make a similar choice except that they use a Brews drain current expression and a model that is similar to the El-Mansy and Boothroyd channel shortening model combined with the Reddi and Sah model. All of the above modified long-channel models include the effect of a field dependent mobility by calculating a constant effective mobility u^* using:

$$u^* = \left[\frac{1}{L} \int_0^L \frac{dy}{u} \right]^{-1} \quad (1.2.1)$$

where (1.2.1) permits the dependence of the mobility on the two components of electric field that are parallel and perpendicular to the silicon surface to be included. The method of averaging shown in (1.2.1) is dictated by the combination of the drain current transport equation and the current continuity equation.

By employing fitting parameters all of the modified long-channel models can give reasonably accurate representations of the behavior of the short-channel MOSFET. To avoid the use of fitting parameters the traditional approach has been to use numerical models that employ finite element or finite difference techniques. The difficulty with this approach is that it has very large computational requirements. "Numerical models" can be divided into three types: those that solve only a two-dimensional (or three-dimensional) Poisson's equation; those that solve a two-dimensional Poisson's equation and a one-dimensional current continuity equation; and those that solve

a two-dimensional Poisson's equation and a two-dimensional current continuity equation. These models are also distinguished by whether they use the finite element or the finite difference technique. But this difference will be neglected here as it has no bearing on the physical nature of the model, only on the details of the program's code.

It is permissible to model the short-channel MOSFET by solving only Poisson's equation and neglecting the current continuity equation if the amount of mobile charge present is small and hence has a negligible effect on the electrostatic potential, as is the case for subthreshold operation. It has also been demonstrated [26] that the drain current in the linear region can be calculated solely from the solution of Poisson's equation, although the effect of mobile charge on the potential must be included in this solution. Examples of MOSFET models based solely upon a two-dimensional solution of Poisson's equation are: Liu, Hoefflinger and Pederson [25] whose program is called TWIST; Greenfield and Dutton [26] whose program is called GEMINI; Scott and Chamberlain [27]; and Viswanathan et al. [28]. A model that uses a three-dimensional solution of Poisson's equation has also been created by Husain and Chamberlain called WAYMOS [29]. These models are of interest here because all two-dimensional analytical models prior to this thesis have solved only Poisson's equation. Hence to one extent or another the two-dimensional analytical models have emulated these numerical models.

The second type of numerical model involves the solution of a two-dimensional Poisson's equation and a one-dimensional current continuity equation. These models are valid for all the normal regimes of operation (i.e. subthreshold, triode and saturation) for a surface conduction MOSFET. But they cannot model buried channel devices, or devices experiencing bulk punchthrough or avalanche breakdown. The first model of this type was by Loeb, Andrew and Love [6]. This work was remarkable in that it was the first numerical model of the MOSFET to appear, but it lacked sophistication because it treated the electrostatic and quasi-Fermi potentials as being identical, thereby neglecting diffusion current. Armstrong, Magowan and Ryan [7] simplified the approach of Loeb, Andrew and Love by treating the mobile charge as an infinitely thin sheet on the silicon surface, thereby making Poisson's equation linear, and by using the depletion approximation. This model was faster but still neglected diffusion current. De La Moneda [8] was able to include diffusion current effects by adapting the algorithm Kim and Yang [31] developed for modelling the JFET, to a form suitable for the MOSFET. This algorithm was originated by Gummel [30], who used it for modelling the bipolar transistor, and hence, will hereafter be referred to as Gummel's algorithm. The essential idea is to analytically solve the one-dimensional current continuity equation for the quasi-Fermi potential as a function of the electrostatic potential. Thus Poisson's equation and the current continuity equation can be solved iteratively

rather than simultaneously. Wilson and Blue [9] improved upon De La Moneda's model by using curved underdiffusions and more sophisticated numerical techniques. The relevance of these numerical models is that in the present thesis a two-dimensional Poisson's equation and one-dimensional current continuity will be solved for the same problem using analytical techniques.

The third type of numerical model solves a two-dimensional Poisson's equation and a two-dimensional current continuity equation. Examples are the model of Vandorpe, Borel, Merckel and Saintot [32], the MINIMOS program by Selberherr, Schütz and Potzl [33,34], the CADDET program by Toyabe, Yamaguchi, Asai and Mock [35] and the PISCES program by Pinto, Rafferty and Dutton [48]. These models are entirely general and valid in all regimes of operation, but the price paid for generality is that they require large amounts of CPU time (typically more than one minute and up to several hours of mainframe CPU time) and can have severe convergence problems. It has not been found possible to date to solve the full two-dimensional problem by analytical techniques.

The third general classification of MOSFET models, which is the subject of this thesis, is two-dimensional analytical models. These models are based upon an approximate two-dimensional analytical solution of Poisson's equation for the MOSFET geometry with approximate boundary conditions. The solution may or may not involve the one-dimensional current continuity equation. In the literature such models have been used to calculate short-channel threshold voltages and subthreshold current, and in

this thesis will be used to calculate above threshold currents too.

The threshold voltage model of Toyabe and Asai [36] was the first published model that fits into this category. They approximated the two-dimensional potential as a family of cubic polynomials thereby reducing the two-dimensional Poisson's equation to a one-dimensional ordinary differential equation. Upon solving this equation a modified form of the long-channel threshold voltage expression was obtained. Several of the parameters in the threshold voltage expression were then used as fitting parameters which were determined using the numerical model CADDET. Toyabe and Asai's model established a feature which is common to all two-dimensional analytical models, this is that the two-dimensional Poisson's equation is solved in a rectangular box or subdomain directly below the channel (see Figure 1.2.1). The accuracy of the model depends upon how accurately Poisson's equation itself is solved and how accurately the boundary conditions used along the sides of the box are specified. In Toyabe and Asai's model the potential and electric field along the bottom boundary are set to zero, or in other words the bottom boundary is treated as the depletion edge which is then of constant depth. This is not a very accurate boundary condition as noted by Ratnakumar and Meindl [12], who found setting the potential to zero along the bottom edge of the box is too restrictive. Furthermore, the depletion edge in a short-channel MOSFET is not constant. It is for this reason that

Figure 1.2.1. The MOSFET geometry with the rectangular domain for the subthreshold solution indicated by cross-hatching.

Tang, Zhang and Navon [37] modified the Toyabe and Asai model to have a depletion depth that varies parabolically with position along the surface.

Toyabe and Asai's threshold voltage model was improved upon by Ratnakumar and Meindl [12] who solved the two-dimensional Poisson's equation exactly using the method of Laplace reduction for an approximate set of boundary conditions. The method of Laplace reduction involves breaking the Poisson problem (i.e. Poisson's equation and its boundary conditions) into several separate problems using one or more trial functions which reduce Poisson's equation to its homogeneous form known as Laplace's equation. The sum of the trial functions and the solutions of the Laplace problems form the solution of the Poisson problem. This method has become the standard technique for two-dimensional analytical MOSFET modelling and is dealt with extensively in this thesis. Although Ratnakumar and Meindl used this sophisticated technique for solving Poisson's equation and identified that setting the electric field to zero along the bottom of the box was a reasonable boundary condition, their other boundary conditions were not very sophisticated as they assumed a constant surface potential and constant potentials on the sides of the box. These side boundary conditions are the equivalent of assuming that the junctions are infinitely deep which is not an appropriate assumption for modern small geometry MOSFETs [43]. Furthermore, in order to obtain a single threshold voltage expression Ratnakumar and Meindl neglected all but the first term in the Fourier series solution of the Laplace problem,

and approximated the hyperbolic sines in this term as exponentials. The resulting threshold voltage expression is indeed simple and gives a first order estimate of the dependence of the threshold voltage reduction on applied biases and device parameters, but in practice it has been found to be quite inaccurate [47] despite Ratnakumar and Meindl's claims to the contrary. Nonetheless, it is a useful expression for circuit simulation when used with fitting parameters [16].

Ratnakumar and Meindl derived their threshold voltage model for devices with a uniformly doped substrate, and for devices with a surface implant that was modelled as an abrupt step. When they solved Poisson's equation in the implanted region the electric field was set to zero along the bottom edge of the implant, but for the uniform substrate solution the electric field was set to zero along a horizontal line at the minimum depletion depth. This inconsistency was rectified by Mehta and Muralidharan [38] who solved two Poisson's equations, one in the surface implant region and one in the bulk region and then matched them, thereby obtaining a more accurate surface implanted threshold voltage model.

The constant surface potential used in Ratnakumar and Meindl's Poisson's equation solution was improved by Poole and Kwong [39] who used a mixed boundary condition based upon the continuity of normal displacement along the silicon-silicon dioxide interface. This boundary condition has the advantages that it is physically accurate, and a charge sheet can easily be incorporated into it for above threshold simulations. Poole and

Kwong used this Poisson's equation solution to derive a short-channel threshold voltage model [39] and a short-channel subthreshold current model [40]. The subthreshold current model uses a very simple expression for the drain current that assumes the inversion layer has a finite thickness related to the electric field normal to the surface. A more accurate analysis would use the drain current expression derived by De La Moneda [8] which will be addressed later in this thesis.

Both the Ratnakumar and Meindl, and the Poole and Kwong models assume infinitely deep junctions thereby overestimating the short-channel effects [43]. Furthermore, neither one can deal with an arbitrary substrate doping profile which is important for modelling modern small geometry MOSFETs that can have highly nonuniform substrates. An alternative approach that overcomes these problems was proposed by Skotnicki [41,42]. He approximated the potential distribution as a family of parabolas, thereby reducing the two-dimensional Poisson's equation to a soluble ordinary differential equation. His solution includes the effect of curved source and drain junctions of finite depth by using parabolic boundary conditions on the sides of the box, and it is valid for an arbitrary substrate doping profile. Its drawbacks are that it is not a true solution of Poisson's equation because of the parabolic assumption and the solution must be modified by a transformation for it to be used for nonzero drain voltages [42].

The threshold voltage model of Kendall and Boothroyd [43] (derived as part of the present research, and reported in

Chapters 2 and 3 of this thesis) has the same features as the Skotnicki model in that it includes the effect of curved source and drain junctions of finite depth and can handle arbitrary substrate doping profiles, but it does not have the disadvantages of Skotnicki's model as it solves Poisson's equation by the Laplace reduction technique. The model makes use of Skotnicki's parabolic side boundary conditions modified by the addition of a trial function that depends upon the nonuniform part of the substrate doping profile. Thus the effect of curved junctions of finite depth is included in the model. These boundary conditions were found to be realistic by comparison to MINIMOS simulations, and by virtue of a cancellation property facilitated a Poisson's equation solution and threshold voltage model that are of the same order of complexity as Poole and Kwong's model, but are valid for an arbitrary substrate doping profile.

An alternative approach to improving upon Ratnakumar and Meindl's constant surface potential boundary condition is to scale the gate oxide thickness by the ratio of the dielectric constants of silicon and silicon dioxide, thereby creating an effective silicon layer with the gate electrode on top of it. The top of the box within which Poisson's equation is solved then advances up to the top of the gate oxide, and the silicon-silicon dioxide interface effectively disappears. This approach was originated by Pfiester, Shott and Meindl [44], and was then used by Nguyen [45], and Marash and Dutton [46]. The advantage of this technique is that the eigenvalues used in the Laplace solutions can be calculated from a closed form equation while the

eigenvalues that result from Poole and Kwong's method [39] must be calculated from an easily solved but still transcendental equation. However, the disadvantages of the oxide scaling technique far outweigh the advantages. First of all to be truly scaled the oxide thickness would have to be increased in the direction of each field line and since it is increased only in the direction normal to the surface this assumes the field lines are all normal to the surface. This is the same approximation used in Poole and Kwong's boundary condition so Pfiester et al.'s technique is no more sophisticated. Secondly, since the silicon-silicon dioxide interface has now effectively vanished it is impossible to include a sheet of charge on the silicon surface. Thus the oxide scaling technique cannot be used above threshold. Thirdly, a two-dimensional solution of Laplace's equation is now required in the gate oxide which seems rather useless since the oxide scaling technique assumes the fields in the gate oxide are one-dimensional. For these three reasons the oxide scaling technique is of doubtful value.

The three papers that used the oxide scaling technique also made improvements to the side boundary conditions: Pfiester, Shott and Meindl continued to use a constant potential down to the junction depth, as Ratnakumar and Meindl did, but from the junction depth to the depletion edge a parabolic boundary condition was used. Nguyen improved upon this by calculating the surface potential using cylindrical junction theory where the doping profile in the source and drain was assumed to be Gaussian, and then allowing the potential to fall off.

parabolically to the depletion edge. This is similar to the boundary condition used by Skotnicki [41] and Kendall and Boothroyd [43]. Marash and Dutton produced even more sophisticated side boundary conditions. They used the same parabolas as Nguyen but added an extra function to the drain side boundary condition in saturation to account for the fact that the gate pulls down the potential near the surface, thereby reversing the sense of the electric field normal to the surface. Both Nguyen, and Marash and Dutton dispensed with Ratnakumar and Meindl's zero field condition along the bottom of the box, and instead used two parabolic potentials that vanished into a central region at zero potential. The side boundary conditions of Marash and Dutton are physically realistic, but it appears that the numerical simulator PISCES [48] must be used to obtain the extra function that is added to the side boundary condition. This is not a satisfactory situation because if a computationally expensive numerical simulation must be done each time the two-dimensional analytical model is used, then the advantage of an analytical approach is lost. What is needed is a two-dimensional analytical model expressed in terms of physically realistic side boundary conditions that do not rely on information from a numerical simulator each time the analytical model is employed. This issue is dealt with later in the thesis.

As well as the subthreshold work, Nguyen [45] also made an attempt at solving Poisson's equation above threshold in the linear part of the triode region. This is the only two-dimensional analytical solution of Poisson's equation

undertaken above threshold before the present thesis. To effect a solution Nguyen neglected the current continuity equation and approximated the surface potential between the source and drain as a linear function. Such a surface potential is a mild approximation in a potential sense, but would produce carrier densities that are grossly in error because of the exponential dependence of carrier density on potential. This is probably the reason Nguyen did not calculate drain currents from the resulting two-dimensional potential as the errors would have been very large. The solution to this problem is to simultaneously solve Poisson's equation and the current continuity equation as is done in the present thesis.

1.3 Objectives of the Thesis

The basic requirement of any practical two-dimensional analytical MOSFET model is that it must be physically realistic enough so that it can simulate the behavior of a short-channel MOSFET without the use of fitting parameters. If this is so then the analytical model can replace numerical models in applications like process and device design, and expand the use of two-dimensional modelling into previously impractical areas such as circuit simulation and statistical modelling. For an analytical model to be physically realistic it must solve Poisson's equation, and the current continuity equation if necessary, exactly or at least very accurately. Furthermore, it must use very realistic physical boundary conditions that do not require the extraction of parameters from numerical simulations.

In addition, it would be an advantage if the model were to have the capability to deal with a field dependent mobility and a substrate doping of arbitrary profile, both of which are important for modelling short-channel MOSFETs. The objectives of this thesis are to produce two models that meet these specifications; one that is valid in subthreshold and another general model that is valid in the subthreshold, triode, and saturation regions.

With reference to the first model, excluding the model of Kendall and Boothroyd [43], which is a part of this thesis, the only subthreshold model that is physically realistic enough to meet these specifications is that of Marash and Dutton, but since its boundary conditions depend upon numerical simulations it must also be disqualified. Even the threshold voltage model of Kendall and Boothroyd will only meet the above requirements for small drain voltages, because its drain side boundary condition is not sophisticated enough to deal precisely with large drain voltages. To meet the specifications the subthreshold model must use an exact analytical solution of Poisson's equation and highly accurate boundary conditions. The most accurate boundary conditions available that are analytically tractable and do not require support from numerical simulators, are the surface continuity of normal displacement condition originated by Poole and Kwong [39], the zero field boundary condition on the bottom of the box proposed by Ratnakumar and Meindl [12], and cubic polynomial side boundary conditions which are an improved form of the Kendall and Boothroyd [43] conditions. The objective is to

solve Poisson's equation exactly using these approximate boundary conditions for a MOSFET with a nonuniformly doped substrate, and from this solution obtain a short-channel threshold voltage model and a short-channel subthreshold current model. In order to calculate accurate subthreshold currents the drain current expression derived by De La Moneda [8] as part of Gummel's algorithm is used. This drain current expression permits the use of a field dependent mobility. Once threshold voltages and subthreshold drain current characteristics are obtained they are tested against numerically simulated and measured data to prove the validity of the model.

The second major objective of the thesis is to develop a general model of the short-channel MOSFET that is valid throughout its entire range of normal operation. This is a very challenging goal because it requires a simultaneous analytical solution of a two-dimensional Poisson's equation and a one-dimensional current continuity equation, which has not been done before. In fact, the prevailing opinion in the modelling community is that such a solution is impossible as demonstrated by the following statement by De La Moneda [8]:

To obtain general and exact solutions for the MOST (i.e. MOSFET) problem, the basic equations must describe the two-dimensional characteristics of the device. This requires integration of a nonlinear system involving the majority-carrier transport equation (i.e. current continuity equation) and Poisson's equation for which exact analytical solutions are improbable.

It is shown in this thesis that such a solution can be obtained using the Laplace reduction method in conjunction with a

sophisticated use of Fourier transform techniques, although it must be conceded to De La Moneda's point of view that this analytical solution still requires numerical integration and iterative techniques for its implementation. The analytical solution assumes that: the inversion layer is modelled as a sheet of charge and hence the current continuity equation becomes one-dimensional; the mobility is constant; and the substrate is uniformly doped. These approximations were necessary to sufficiently simplify the problem so that there was a reasonable chance of finding an analytical solution. Once the solution that assumes a constant mobility and uniform substrate is found, it provides a basis to derive a solution that includes the effects of a field dependent mobility and a nonuniform substrate, both of which are necessary to model actual short-channel devices. The techniques that could be used to do this are outlined, but not completed, in Chapter 8. This is because it was felt that within the scope of the Ph.D. thesis it was sufficient to demonstrate the viability of the analytical technique for a constant mobility and uniform substrate, and then outline how the solution could be extended to include a field dependent mobility and a nonuniform substrate.

The analytical solution of a two-dimensional Poisson's equation and a one-dimensional current continuity equation utilizes the surface boundary condition of Poole and Kwong, the zero field bottom boundary condition of Ratnakumar and Meindl, and a new type of side boundary condition that is physically accurate under all bias conditions and can be used independently

of numerical simulations. It is shown that the simultaneous analytical solution produces a closed form equation for the drain current which is valid in all the normal regimes of operation, and that when the solution is implemented in software it runs 5 to 300 times faster than even the fastest comparable numerical model. The drain current characteristics are compared to numerically simulated data to demonstrate the validity and accuracy of the general model.

1.4 Outline of the Thesis

The two-dimensional analytical solution of Poisson's equation for a MOSFET with a nonuniformly doped substrate, and its associated threshold voltage and subthreshold current models, are derived in Chapter 2. In Chapter 3 these models are tested and verified by comparison to measured and numerically simulated data. This includes a description of the detailed characterization of the devices used for the measurements. The simultaneous solution of a two-dimensional Poisson's equation and a one-dimensional current continuity equation for the short-channel MOSFET with a uniform substrate is presented in Chapter 4. The two methods that were used to complete the solution are the iterative function technique and the Fourier analysis technique. These are described in Chapters 5 and 6 respectively. At present, the use of the latter technique is limited to the case of a constant mobility. The general MOSFET model is verified by comparison with the results of MINIMOS simulations in Chapter 7. The general model is the first model

of its type and because of this it has opened up a whole range of new possibilities for further research. These are discussed in Chapter 8 and include the extension of the general model to include the effects of a field dependent mobility and a nonuniform substrate. Finally, the conclusions of the thesis are presented in Chapter 9.

CHAPTER 2

THE SUBTHRESHOLD MODEL

2.1 The Poisson Problem and Approximating the Boundary Conditions

2.1a Formulating the Problem

To develop a model of the behavior of the short-channel MOSFET in subthreshold a Poisson problem must be solved. This consists of the solution of a two-dimensional Poisson's equation for the geometry and boundary conditions of the MOSFET. The most important aspect in making such a solution accurate is to employ as few approximations as possible in the boundary conditions and in the method of solution. For a two-dimensional analytical solution this means using the method of Laplace reduction to solve Poisson's equation, because it results in an exact solution, and using the most accurate boundary conditions on the four sides of the rectangular box (see Figure 1.2.1) that will still permit an analytically tractable solution.

The complete form of Poisson's equation in a semiconductor such as silicon is known as the Shockley-Poisson equation which in two dimensions has the form

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{q}{\epsilon_s} [n_i e^{\beta(\psi - \phi_n)} - n_i e^{\beta(\phi_p - \psi)} + N_A - N_D] \quad (2.1.1)$$

where x and y are defined by the coordinate system in Figure 1.2.1, ψ is the electrostatic potential, ϕ_n is the

electron quasi-Fermi potential, ϕ_p is the hole quasi-Fermi potential, N_A is the concentration of ionized acceptor atoms, N_D is the concentration of ionized donor atoms, n_i is the intrinsic carrier concentration, q is the charge on an electron, ϵ_s is the permittivity of silicon, and β is $q/(kT)$ where k is Boltzmann's constant and T is the absolute temperature. The electrostatic and two quasi-Fermi potentials are referenced to a zero electrostatic potential in the neutral bulk. This is equivalent to referencing the potentials to the middle of the band gap. Equation (2.1.1) uses Boltzmann rather than Fermi-Dirac statistics and hence is only valid under nondegenerate conditions. The four terms on the right hand side of (2.1.1) are respectively the electron, hole, acceptor and donor concentration terms. The Shockley-Poisson equation is highly nonlinear due to its exponential dependence on ψ making it a very difficult equation to solve. However, in subthreshold it can be simplified by applying the depletion approximation from the silicon surface down to the depletion region edge, thereby eliminating the electron and hole terms. The depletion approximation assumes that the depletion region is devoid of carriers of either type right up to an abrupt depletion region edge after which point the carriers are present again and make the substrate neutral. It is applicable in the vicinity of the silicon surface because in subthreshold the concentration of carriers in the inversion layer is negligible compared to the concentration of ionized dopant atoms. An additional simplification to (2.1.1) is possible because the processes that are used to fabricate MOSFETs

generally only introduce nonuniformities in the substrate doping in the x direction, so it is permissible to replace $N_A - N_D$ with the substrate doping profile $N(x)$. Thus in the depletion region (2.1.1) becomes:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{qN(x)}{\epsilon_s} \quad (2.1.2)$$

Poisson's equation in this form is linear making it analytically tractable. If the substrate doping profile $N(x)$ is divided into a constant substrate doping N_B and a nonuniform doping $N_n(x)$, then (2.1.2) becomes

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{qN_B}{\epsilon_s} + \frac{qN_n(x)}{\epsilon_s} = K + f(x) \quad (2.1.3)$$

The linearity of (2.1.3) can then be taken advantage of to divide it into three Poisson's equations:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x) \quad (2.1.4a)$$

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = K \quad (2.1.4b)$$

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad (2.1.4c)$$

$$\text{where } \psi(x, y) = T(x) + U(x) + V(x, y) \quad (2.1.5)$$

$T(x)$ and $U(x)$ are respectively the nonuniformity and uniformity trial functions that are used to reduce Poisson's equation to its homogeneous form called Laplace's equation given by (2.1.4c), whose solution is $V(x,y)$.

2.1b Boundary Conditions

The solutions for T , U and V will be presented in the next section, but it is first necessary to establish the boundary conditions on the four sides of the rectangular box so that the description of the Poisson problem may be completed. The boundary condition used along the silicon surface on the top of the box is based upon the continuity of normal electric displacement and its use was first proposed by Poole and Kwong [39,40] who expressed it as

$$\psi(0,y) - R t_{ox} \frac{\partial \psi}{\partial x}(0,y) = V_{GB}' \quad (2.1.6)$$

where t_{ox} is the gate oxide thickness, V_{GB}' is the gate potential with respect to the flat band voltage V_{FB} (i.e. $V_{GB}' = V_{GB} - V_{FB}$), and R is the ratio of the dielectric constants of silicon and silicon dioxide ϵ_s/ϵ_{ox} . This boundary condition is very realistic because the continuity of normal displacement is a basic physical constraint, but it is still approximate as it assumes a one-dimensional solution of Laplace's equation in the gate oxide and neglects the presence of electron charge at the silicon surface. For a MOSFET with a thin gate oxide that is

operated in subthreshold, which is the case under consideration here, both these approximations are valid.

The boundary condition used along the bottom of the box is that the electric displacement crossing this boundary is zero as recommended by Ratnakumar and Meindl [12]:

$$\frac{\partial \psi}{\partial x}(d, y) = 0 \quad (2.1.7)$$

where d is the minimum depletion depth (see Figure 1.2.1). This boundary condition is only exact at the intersection point of the line $x = d$ and the depletion edge. At all other points it neglects the electric displacement crossing the line $x = d$ with the largest error being near the sides of the box. The reason this is acceptable is that the total neglected displacement crossing this boundary is small compared to the total error in the displacement crossing the other three boundaries which is also small. The errors on the other three boundaries are caused by the approximate nature of the boundary conditions used. Furthermore, the region of practical interest is near the surface so the boundary condition used on the bottom of the box is not as critical as for the other three sides. An advantage of using this boundary condition is that it makes the analytical solution of Poisson's equation particularly simple.

What is required for the boundary conditions on the sides of the box are functions that accurately represent the potential variations along the lines $y = 0$ and $y = L$ from the surface down to the depletion edge, where L is the channel length. These

boundary conditions are very critical because they control the amount of electric displacement entering the sides of box from the source and drain, which is what causes short-channel effects. Polynomials are a particularly appropriate set of functions to choose for the side boundary conditions because their coefficients can be chosen to satisfy known physical conditions and they make the Fourier analysis required for the solution of $V(x,y)$ fast and easy. Skotnicki [41,42] originated this approach by choosing parabolas for the side boundary conditions of the form:

$$\psi(x,0) = \psi(0,0) (1 - x/x_S)^2 \quad (2.1.8)$$

where x_S is the depletion depth at $y = 0$ (see Figure 1.2.1) calculated from an approximation of the underdiffusion as an ellipse, and $\psi(0,0)$ is calculated by an exponential-constant approximation to the source-substrate junction doping profile following the work of Chawla and Gummel [49].

Skotnicki reported that (2.1.8) was a reasonable approximation to the potential variation predicted by MINIMOS [33] along the line $y = 0$. This was confirmed by the author for two equations similar to (2.1.8) namely

$$\psi(x,0) = V_{MS} (1 - x/x_S)^2 \quad (2.1.9)$$

$$\psi(x,L) = V_{MD} (1 - x/x_D)^2 \quad (2.1.10)$$

$$\text{where } V_{MS} = V_S - V_C \quad (2.1.11)$$

$$V_{MD} = V_D - V_C \quad (2.1.12)$$

$$V_S = \psi_{bi} + V_{SB} \quad (2.1.13)$$

$$V_D = \psi_{bi} + V_{SB} + V_{DS} \quad (2.1.14)$$

$$x_S = \sqrt{R_S^2 - r_j^2} \quad (2.1.15)$$

$$x_D = \sqrt{R_D^2 - r_j^2} \quad (2.1.16)$$

where V_S and V_D are the potentials at the source and drain contacts respectively, ψ_{bi} is the built-in junction potential, V_{SB} and V_{DS} are the source-substrate and drain-source applied potentials respectively, x_S and x_D are the depletion depths at $y = 0$ and $y = L$, R_S and R_D are the source and drain depletion layer radii, and r_j is the junction depth for an assumed quarter-cylindrical metallurgical junction configuration (see Figure 1.2.1). The method for calculating R_S and R_D will be discussed shortly. The parameter V_C accounts for the voltage drop between the source contact and the point $(0,0)$, and the voltage drop between the drain contact and the point $(0,L)$. The value of V_C is held constant at 0.15 V for simplicity, since this is the average voltage drop observed in MINIMOS simulations. A constant V_C provides an alternative to the complicated Chawla and Gummel method used by Skotnicki. This method is based upon calculating the potential at the metallurgical junction of a planar pn junction where the source (or drain) has an exponential doping profile and the substrate has a constant doping profile. The Chawla and Gummel method is not appropriate to calculate the potential at points $(0,0)$ and $(0,L)$ because their theory applies

to planar pn junctions, and what is being considered here is a cylindrical junction that abruptly ends at the silicon surface and is heavily influenced by the gate. Thus using a constant V_C to calculate the potential at points $(0,0)$ and $(0,L)$ is no more physically justifiable than using the Chawla and Gummel method, but it is employed because it is simple and produces reasonably accurate results.

The validity of (2.1.9) and (2.1.10) was checked by comparing them to potential profiles extracted from MINIMOS simulations. The results of such a comparison can be seen in Figure 2.1.1 for $V_{GB} = 0$ and 5 V for a uniform substrate device whose parameters are given in Table 2.1.1. Note that the parameter N_S listed in the table is the peak doping in the source and drain regions, and W is the width of the device. The plots in Figure 2.1.1 are all cut off at depth d because this is the extent of the boundary conditions that are required for the Poisson solution. It is clear from the four plots in this figure that (2.1.9) and (2.1.10) provide a reasonable estimate of the potential profile along the sides of the box for this device.

The sort of accuracy shown in Figure 2.1.1 was typical for uniform substrate devices but it was found that for devices with nonuniformly doped substrates it was necessary to include the effect of the nonuniformity in the side boundary conditions. This was done with the aid of the following thought experiment: Imagine that the potential profile, say $\psi(x,0)$, is a parabola for a device with a uniform substrate. Dopant atoms are then added a few at a time to create a nonuniformity, say a shallow surface

The quality of this microfiche is heavily dependent upon the quality of the thesis submitted for microfilming.

Please refer to the National Library of Canada target (sheet 1, frame 2) entitled:

CANADIAN THESES

NOTICE

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage.

Veuillez consulter la cible de la Bibliothèque nationale du Canada (microfiche 1, image 2) intitulée:

THÈSES CANADIENNES

AVIS

(a)

(b)

(c)

(d)

Figure 2.1.1. Comparison of the quadratic side boundary conditions to the results of MINIMOS simulations for a uniform substrate device where for (a) and (b) $L = 0.5 \mu\text{m}$, $V_{GS} = 0.472 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$, and for (c) and (d) $L = 1.5 \mu\text{m}$, $V_{GS} = 1.380 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 5 \text{ V}$.

Table 2.1.1. Device parameters for the uniform substrate, NMOS and CMOS devices.

Device Type	N_B (cm^{-3})	N_S (cm^{-3})	t_{ox} ($\text{\AA}$)	r_j (μm)	V_{FB} (V)	W (μm)
Uniform Substrate	5.0×10^{15}	2.3×10^{20}	500	0.5	-0.5	24
NMOS	2.0×10^{15}	2.0×10^{20}	500	1.0	-1.0	24
CMOS	1.61×10^{16}	2.3×10^{20}	500	0.5	-0.72	24

$$\text{NMOS: } N(x) = 3 \times 10^{16} e^{-(x/0.1)^2} + 2 \times 10^{15} (\text{cm}^{-3})$$

$$\text{CMOS: } N(x) = 2.5 \times 10^{16} e^{-(x/2.2)^2} - 8 \times 10^{15} e^{-(x/0.13)^2} - 5.9 \times 10^{15} (\text{cm}^{-3})$$

where x is in microns

implant to modify the threshold voltage. The charge on the ionized dopant atoms then seeks to modify the parabolic potential profile and because Poisson's equation is linear in a depletion region the potential profile will be modified in an additive fashion. But the source has a firm hold on the potential at the surface and does not let it change. The resulting potential profile must therefore depend linearly on the trial function $T(x)$ which results from solving Poisson's equation for the nonuniformity in the doping profile (see equation (2.1.4a)), but maintains the same surface potential. A formulation that satisfies these requirements is

$$\psi(x,0) = [V_{MS} - T(0)](1 - x/x_S)^2 + T(x) \quad (2.1.17)$$

$$\psi(x,L) = [V_{MD} - T(0)](1 - x/x_D)^2 + T(x) \quad (2.1.18)$$

where $T(x)$ is determined for a specific nonuniform doping profile by integrating (2.1.4a).

These boundary conditions are compared to the results of MINIMOS simulations for a typical device for an NMOS process in Figure 2.1.2, and for a typical n-channel device fabricated in the p-well of a CMOS process in Figure 2.1.3. The device parameters for these devices are given in Table 2.1.1. Note that the value of N_B given for the CMOS device is an averaged value for the first micron of the p-well. The substrate doping profiles of the NMOS and CMOS devices are shown in Figures 2.1.4 and 2.1.5 respectively. It is clear from Figures 2.1.2 and 2.1.3 that the accuracy of the boundary conditions is quite good for

(a)

(b)

(c)

(d)

Figure 2.1.2. Comparison of the quadratic side boundary conditions to the results of MINIMOS simulations for an NMOS device where for (a) and (b) $L = 1.2 \mu\text{m}$, $V_{GS} = 0.383 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$, and for (c) and (d) $L = 1.5 \mu\text{m}$, $V_{GS} = 0.713 \text{ V}$, $V_{DS} = 0.1 \text{ V}$ and $V_{SB} = 5 \text{ V}$.

(a)

(b)

(c)

(d)

Figure 2.1.3. Comparison of the quadratic side boundary conditions to the results of MINIMOS simulations for a CMOS device where for (a) and (b), $L = 0.5 \mu\text{m}$, $V_{GS} = 0.697 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$, and for (c) and (d) $L = 0.8 \mu\text{m}$, $V_{GS} = 2.004 \text{ V}$, $V_{DS} = 0.1 \text{ V}$ and $V_{SB} = 5 \text{ V}$.

Figure 2.1.4. Substrate doping profile for the NMOS device.

Figure 2.1.5. Substrate doping profile for the CMOS device.

the CMOS device but rather poor for the NMOS device. This is because the nonuniformities in the substrate doping profile of the CMOS device are not very pronounced but those of the NMOS device are. Thus it appears that to get accuracy for the NMOS device (2.1.17) and (2.1.18) need to be modified even further.

2.1c Improved Side Boundary Conditions

Thus far, the side boundary conditions have utilized information about the potential and field at the depletion edge and the potential at the surface, but information about the electric field at the surface provided by the condition of the continuity of normal displacement has not been employed. Adding the surface field to the three conditions already used gives enough information to determine the coefficients of a cubic polynomial, and given the linearity of Poisson's equation it would seem natural to add the cubic polynomial to the expressions given by (2.1.17) and (2.1.18). If the cubic polynomial is chosen to give the correct surface field, and so that it does not disturb the potential and electric field at the depletion edge or the potential at the surface, then the four conditions needed to calculate the cubic polynomial along the line $y = L$ (the same approach is applicable for the source boundary) are:

$$\psi(0) = 0 \quad (2.1.19)$$

$$\frac{d\psi}{dx}(0) = F_D \quad (2.1.20)$$

$$\psi(x_D) = 0 \quad (2.1.21)$$

$$\frac{dC}{dx}(x_D) = 0 \quad (2.1.22)$$

where F_D is the correction to the electric field normal to the surface at $(0,L)$ necessary to produce continuity of normal displacement, and $C(x)$ is the cubic polynomial. These four conditions yield the following expression for $C(x)$:

$$C(x) = F_D x_D \left[\left(\frac{x}{x_D} \right) - 2 \left(\frac{x}{x_D} \right)^2 + \left(\frac{x}{x_D} \right)^3 \right] \quad (2.1.23)$$

It has been found that the magnitude of the correction to the surface field F_D must be limited lest it send the potential profile $\psi(x,L)$ above the potential at the drain contact or send it below the potential at the substrate contact, both of which are nonphysical. This is mathematically equivalent to restricting the initial slope of $C(x)$ to lie within two bounds as shown in Figure 2.1.6. When $C(x)$ tries to exceed these bounds, attempts to produce continuity of normal displacement at the silicon surface are abandoned. In subthreshold the mobile charge may be neglected and then the continuity of normal displacement is expressed as:

$$C_{ox} (V_{GB} - V_{MD}) + \epsilon_s T_D = 0 \quad (2.1.24)$$

$$\text{where } T_D = \frac{\partial \psi}{\partial x}(0, L) \quad (2.1.25)$$

where C_{ox} is the capacitance per unit area of the gate oxide. Enforcing the continuity of normal displacement while restricting

Figure 2.1.6. Limiting the slope at $x = 0$ of the cubic polynomial $C(x)$.

the initial slope for $C(x)$ yields the following set of rules for calculating F_D :

$$T_D = \frac{V_{MD} - V_{GB}}{Rt_{ox}} \quad (2.1.26)$$

$$\text{since } Rt_{ox} = \epsilon_s / C_{ox} \quad (2.1.27)$$

Defining the Kronecker deltas δ_{D1} and δ_{D2} as:

$$\text{If } T_D < 0 \text{ then } \delta_{D1} = 1 \quad (2.1.28a)$$

$$\text{If } T_D > 0 \text{ then } \delta_{D1} = 0 \quad (2.1.28b)$$

$$\text{If } T_D \delta_{D1} + \frac{2V_L}{x_D} + F(d) - F(0) \geq 0 \text{ then } \delta_{D2} = 1 \quad (2.1.29a)$$

$$\text{If } T_D \delta_{D1} + \frac{2V_L}{x_D} + F(d) - F(0) < 0 \text{ then } \delta_{D2} = 0 \quad (2.1.29b)$$

allows F_D to be written as

$$F_D = [T_D \delta_{D1} + \frac{2V_L}{x_D} + F(d) - F(0)] \delta_{D2} \quad (2.1.30)$$

$$\text{where } V_L = V_{MS} - T(0) \quad (2.1.31)$$

$$F(x) = \int f(x) dx \quad (2.1.32)$$

$$f(x) = \frac{qN_n(x)}{\epsilon_s} \quad (\text{see eqn. (2.1.3)}) \quad (2.1.33)$$

Note that $F(x)$ can be calculated in closed form if $N_n(x)$ is expressed as the sum of Gaussians plus a constant, which is common. Similarly for F_S in the case of the source boundary:

$$T_S = \frac{V_{MS} - V_{GB}}{Rt_{ox}} \quad (2.1.34)$$

$$\text{If } T_S \leq 0 \text{ then } \delta_{S1} = 1 \quad (2.1.35a)$$

$$\text{If } T_S > 0 \text{ then } \delta_{S1} = 0 \quad (2.1.35b)$$

$$\text{If } T_S \delta_{S1} + \frac{2V_0}{x_S} + F(d) - F(0) > 0 \text{ then } \delta_{S2} = 1 \quad (2.1.36a)$$

$$\text{If } T_S \delta_{S1} + \frac{2V_0}{x_S} + F(d) - F(0) < 0 \text{ then } \delta_{S2} = 0 \quad (2.1.36b)$$

$$F_S = [T_S \delta_{S1} + \frac{2V_0}{x_S} + F(d) - F(0)] \delta_{S2} \quad (2.1.37)$$

$$\text{where } V_0 = V_{MS} - T(0) \quad (2.1.38)$$

The cubic polynomial $C(x)$ can now be added to (2.1.17) and (2.1.18) to obtain the complete cubic polynomial side boundary conditions:

$$\psi(x, 0) = V_0 \left(1 - \frac{x}{x_S} \right)^2 + T(x) + F_S x_S \left(\frac{x}{x_S} + 2 \left(\frac{x}{x_S} \right)^2 + \left(\frac{x}{x_S} \right)^3 \right) \quad (2.1.39)$$

$$\psi(x, L) = V_L \left(1 - \frac{x}{x_D}\right)^2 + T(x) + F_D x_D \left[\frac{x}{x_D} - 2 \left(\frac{x}{x_D}\right)^2 + \left(\frac{x}{x_D}\right)^3 \right] \quad (2.1.40)$$

Note that depending upon the value of the Kronecker deltas δ_{S1} , δ_{S2} , δ_{D1} and δ_{D2} all or part of F_S and F_D can vanish. In fact, if δ_{S2} and δ_{D2} are zero then (2.1.39) and (2.1.40) reduce to the quadratic boundary conditions given by (2.1.17) and (2.1.18), and if $T(x)$ is set to zero these boundary conditions reduce to those given by (2.1.9) and (2.1.10). Furthermore, the infinitely deep junction boundary conditions used by Ratnakumar and Meindl [12] and Poole and Kwong [39,40] may be obtained by setting V_C to zero and x_S and x_D to infinity in (2.1.9) and (2.1.10). So the four sets of boundary conditions form a hierarchy where additional complexity brings additional accuracy. It should be noted, however, that this is not the author's last word on the side boundary conditions as an even more physically and mathematically sophisticated set will be developed in Chapter 4 for use with the general model. They were not used with the subthreshold model because they were developed after this work was finished.

The accuracy of (2.1.39) and (2.1.40) was tested by comparing them to potential profiles extracted from MINIMOS simulations. The results of this comparison for the uniform substrate, NMOS and CMOS devices can be seen in Figures 2.1.7, 2.1.8 and 2.1.9 respectively. It is obvious when these figures are compared to Figures 2.1.1, 2.1.2 and 2.1.3 that the cubic boundary conditions are in general more accurate than the

(a)

(b)

(c)

(d)

Figure 2.1.7. Comparison of the cubic side boundary conditions to the results of MINIMOS simulations for a uniform substrate device where for (a) and (b) $L = 0.5 \mu\text{m}$, $V_{GS} = 0.472 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$, and for (c) and (d) $L = 1.5 \mu\text{m}$, $V_{GS} = 1.380 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 5 \text{ V}$.

(a)

(b)

(c)

(d)

Figure 2.1.8. Comparison of the cubic side boundary conditions to the results of MINIMOS simulations for an NMOS device where for (a) and (b) $L = 1.2 \mu\text{m}$, $V_{GS} = 0.383 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$, and for (c) and (d) $L = 1.5 \mu\text{m}$, $V_{GS} = 0.713 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 5 \text{ V}$.

(a)

(b)

(c)

(d)

Figure 2.1.9. Comparison of the cubic side boundary conditions to the results of MINIMOS simulations for a CMOS device where for (a) and (b) $L = 0.5 \mu\text{m}$, $V_{GS} = 0.697 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$, and for (c) and (d) $L = 0.8 \mu\text{m}$, $V_{GS} = 2.004 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 5 \text{ V}$.

quadratic boundary conditions especially for the NMOS device. Thus using the cubic boundary conditions in the solution of Poisson's equation will produce a more accurate short-channel MOSFET model. There is also the additional advantage that the quadratic, uniform substrate quadratic, and infinitely deep junction boundary conditions can be obtained as special cases of the cubic boundary conditions.

2.1d Calculating the Cylindrical Depletion Layer Radii

The only remaining task to complete the description of the side boundary conditions is to develop a method to calculate the cylindrical depletion layer radii R_S and R_D so that the depletion depths x_S and x_D can be calculated from equations (2.1.15) and (2.1.16) (see Figure 1.2.1). The radii R_S and R_D are important because they permit the effect of the curved underdiffusions of the source and drain to be included in the solution of Poisson's equation for the rectangular box in a simple way.

To calculate R_S and R_D a two-dimensional, cylindrically symmetric Poisson's equation must be solved in the depletion region near the source and drain underdiffusions. If the depletion edge and metallurgical junction in this region are assumed to be circular arcs (see Figure 2.1.10), and the substrate is uniformly doped, then Poisson's equation may be written in polar coordinates as

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{d\psi}{dr} \right) = \frac{qN_B}{\epsilon_s} \quad (2.1.41)$$

Figure 2.1.10. Depletion layer geometry near the source underdiffusion.

The three boundary conditions necessary to solve (2.1.41) for R_S are

$$\psi(R_S) = 0 \quad (2.1.42)$$

$$\frac{d\psi}{dr}(R_S) = 0 \quad (2.1.43)$$

$$\psi(r_j) = \psi_{bi} + V_{SB} \quad (2.1.44)$$

The first two conditions are the zero potential and zero field conditions at a depletion edge, and the third represents the assumption that the entire source region is at the potential of the source contact where V_{SB} is the reverse bias across the source-substrate junction. Integrating (2.1.41) twice and using the conditions given by (2.1.43) and (2.1.44) yields the following expression for the potential $\psi(r)$.

$$\psi(r) = \psi_{bi} + V_{SB} + M \left[\frac{r^2 - r_j^2}{2} - R_S^2 \ln \left(\frac{r}{r_j} \right) \right] \quad (2.1.45)$$

$$\text{where } M = \frac{qN_B}{2\epsilon_s} \quad (2.1.46)$$

Applying the third condition (2.1.42) to (2.1.45) and rearranging yields the following transcendental equation in R_S [50]:

$$R_S = \left[\frac{\frac{R_S^2 - r_j^2}{2} + \frac{\psi_{bi} + V_{SB}}{M}}{\ln \left(\frac{R_S}{r_j} \right)} \right]^{1/2} \quad (2.1.47)$$

Using an initial guess of

$$R_S = r_j + 0.87 \sqrt{\frac{\psi_{bi} + V_{SB}}{M}} \quad (2.1.48)$$

which is r_j plus 87% of the depletion width for a planar junction, and a single iteration using (2.1.47) gives R_S to an accuracy of 4 or 5 significant digits. This is sufficiently accurate for virtually all applications. Equations (2.1.47) and (2.1.48) may also be used to calculate the drain depletion layer radius R_D simply by replacing R_S with R_D and V_{SB} with $V_{SB} + V_{DS}$.

The calculation of the depletion layer radius becomes much more complicated if the substrate is nonuniformly doped. This could be approached by solving a two-dimensional Poisson's equation in the depletion region but because R_S and R_D are to be used in the approximate side boundary conditions this sort of complication is unwarranted. A better alternative is simply to average the doping profile between the metallurgical junction and the depletion region edge and use this effective substrate doping N_{Be} in (2.1.47) and (2.1.48).

Referring to Figure 2.1.10 the doping can be averaged using

$$N_{Be} = \frac{1}{R_S - r_j} \int_{r_j}^{R_S} N(x) \, dr \quad (2.1.49)$$

Since the doping profile $N(x)$ is a function of the distance from the surface x not the radial distance r the integral must be

transformed. Using the similarity of triangles ABC and ADE (see Figure 2.1.10)

$$\frac{x}{r} = \frac{x_S}{R_S} \quad (2.1.50)$$

and noting that

$$N(x) = N_B + N_n(x) \quad (2.1.51)$$

(2.1.49) may be transformed to

$$N_{Be} = \frac{1}{R_S - r_j} \frac{R_S}{x_S} \int_{\frac{x_S r_j}{R_S}}^{x_S} [N_B + N_n(x)] dx \quad (2.1.52)$$

Integrating (2.1.52) yields

$$N_{Be} = \frac{\epsilon_S}{q} \frac{R_S}{x_S (R_S - r_j)} \left[\frac{K x_S}{R_S} (R_S - r_j) + F(x_S) - F\left(\frac{x_S r_j}{R_S}\right) \right] \quad (2.1.53)$$

$$\text{where } F(x) = \int f(x) dx = \int \frac{q N_n(x)}{\epsilon_S} dx \quad (2.1.54)$$

Using (2.1.53), just adds a few extra steps to the iterative procedure for calculating R_S . After R_S has been estimated using (2.1.48), N_{Be} is calculated from (2.1.53) and used to calculate effective values of ψ_{bi} and M from

$$\psi_{bi} = \frac{kT}{q} \ln \left(\frac{N_S N_{Be}}{n_i^2} \right) \quad (2.1.55)$$

$$M = \frac{qN_{Be}}{2\epsilon_S} \quad (2.1.56)$$

then R_S is calculated from (2.1.47). The process is then repeated and converges rapidly to the correct value of R_S .

Once R_S and R_D are obtained from (2.1.47) and (2.1.48), and if the substrate is nonuniform from (2.1.53), (2.1.55) and (2.1.56), then x_S and x_D may be calculated from (2.1.15) and (2.1.16). The quantities x_S and x_D are then available to complete the description of the side boundary conditions.

2.2 Solution of the Poisson Problem

The Poisson problem that must be solved to obtain a subthreshold model for the short-channel MOSFET consists of Poisson's equation and the boundary conditions on the four sides of the rectangular box under consideration (see Figure 1.2.1). Collecting together the equations given in Section 2.1 this problem is stated as:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{qN_B}{\epsilon_S} + \frac{qN_n(x)}{\epsilon_S} = K + f(x) \quad (2.1.3)$$

where (2.1.3) is Poisson's equation and the boundary conditions are:

$$\psi(0, y) - R t_{ox} \frac{\partial \psi}{\partial x}(0, y) = V_{GB} \quad (2.1.6)$$

$$\frac{\partial \psi}{\partial x}(d, y) = 0 \quad (2.1.7)$$

$$\psi(x, 0) = V_0 \left(1 - \frac{x}{x_S}\right)^2 + T(x) + F_S x_S \left[\frac{x}{x_S} - 2 \left(\frac{x}{x_S}\right)^2 + \left(\frac{x}{x_S}\right)^3\right] \quad (2.1.39)$$

$$\psi(x, L) = V_L \left(1 - \frac{x}{x_D}\right)^2 + T(x) + F_D x_D \left[\frac{x}{x_D} - 2 \left(\frac{x}{x_D}\right)^2 + \left(\frac{x}{x_D}\right)^3\right] \quad (2.1.40)$$

$$\text{where } V_0 = V_S - V_C - T(0) \quad (2.2.1)$$

$$V_L = V_D - V_C - T(0) \quad (2.2.2)$$

As introduced in the previous section, to use the Laplace reduction technique Poisson's equation is broken into three parts

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x) \quad (2.1.4a)$$

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = K \quad (2.1.4b)$$

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad (2.1.4c)$$

$$\text{where } \psi(x, y) = T(x) + U(x) + V(x, y) \quad (2.1.5)$$

$T(x)$ and $U(x)$ are the trial functions associated with the ideal long-channel MOSFET for which short-channel effects are entirely neglected. The solution of Poisson's equation in subthreshold for such a device is $T(x) + U(x)$, and as indicated by the notation, is independent of y . Because of the lack of y

dependence (2.1.4a) and (2.1.4b) reduce to the ordinary differential equations

$$\frac{d^2 T}{dx^2} = f(x) \quad (2.2.3)$$

$$\frac{d^2 U}{dx^2} = K \quad (2.2.4)$$

It should be noted that this simplification involves no approximations. Integrating (2.2.3) twice and applying the boundary conditions

$$T(d) = 0 \quad (2.2.5)$$

$$\frac{dT}{dx}(d) = 0 \quad (2.2.6)$$

where (2.2.6) neglects the small electric field at the depletion edge due to the nonuniform doping profile, then $T(x)$ is given by

$$T(x) = G(x) - G(d) + (d - x) F(d) \quad (2.2.7)$$

where d is the minimum depletion depth in the short-channel MOSFET, and $F(x)$ and $G(x)$ are respectively the first and second indefinite integrals of $f(x)$, i.e.

$$F(x) = \int f(x) dx \quad (2.2.8)$$

$$G(x) = \iint f(x) dx dx \quad (2.2.9)$$

It is usually possible to specify $f(x)$ as the sum of Gaussians plus a constant for which $F(x)$ and $G(x)$ are closed form functions.

$U(x)$ is determined by integrating (2.2.4) twice and applying the boundary conditions

$$U(d) = p \quad (2.2.10)$$

$$\frac{dU}{dx}(d) = 0 \quad (2.2.11)$$

where the parameter p is chosen so that after Laplace reduction (2.1.6) involves only $V(x,y)$ and its first derivative. The parameter p tends to zero as the channel length goes to infinity thus yielding the correct long-channel boundary condition. The integration of (2.2.4) yields

$$U(x) = \frac{K}{2} (d - x)^2 + p \quad (2.2.12)$$

To determine p substitute (2.2.7) and (2.2.12) into (2.1.5), and substitute (2.1.5) into (2.1.6), then choose p so that

$$V(0,y) - R t_{ox} \frac{\partial V}{\partial x}(0,y) = 0 \quad (2.2.13)$$

This yields

$$p = R t_{tox} [F(0) - \frac{1}{2} K d^2 - K d] - T(0) - \frac{K}{2} d^2 + V_{GB} \quad (2.2.14)$$

and hence

$$U(x) = \frac{K}{2} (x^2 - 2dx) + V_{GB} - T(0) + R t_{ox} [F(0) - F(d) - Kd] \quad (2.2.15)$$

The other three boundary conditions for $V(x,y)$ are determined by rearranging (2.1.5) to

$$V(x,y) = \psi(x,y) - T(x) - U(x) \quad (2.2.16)$$

Substituting (2.1.7), (2.1.39), (2.1.40), (2.2.7) and (2.2.12) into (2.2.16), and recalling (2.1.4c) and (2.2.13), permits the Laplace problem for $V(x,y)$ to be written down as:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad (2.1.4c)$$

where (2.1.4c) is Laplace's equation and the boundary conditions are:

$$V(0,y) - R t_{ox} \frac{\partial V}{\partial x}(0,y) = 0 \quad (2.2.13)$$

$$\frac{\partial V}{\partial x}(d,y) = 0 \quad (2.2.17)$$

$$V(x,0) = V_0 \left(1 - \frac{x}{x_S}\right)^2 + F_S x_S \left[\frac{x}{x_S} - 2 \left(\frac{x}{x_S}\right)^2 + \left(\frac{x}{x_S}\right)^3\right]$$

$$- \frac{K}{2} (x^2 - 2dx) - V_{GB} + T(0) - R t_{ox} [F(0) - F(d) - Kd] \quad (2.2.18)$$

$$V(x, L) = V_L \left(1 - \frac{x}{x_D}\right)^2 + F_D x_D \left[\frac{x}{x_D} - 2 \left(\frac{x}{x_D}\right)^2 + \left(\frac{x}{x_D}\right)^3\right]$$

$$- \frac{K}{2} (x^2 - 2dx) - V_{GB} + T(0) - R t_{ox} [F(0) - F(d) - Kd]$$

(2.2.19)

This Laplace problem can be solved by the well known separation of variables technique where it is assumed that $V(x, y)$ has the form

$$V(x, y) = \sum_{i=0}^{\infty} X_i(x) Y_i(y) \quad (2.2.20)$$

Using this form it is found that for the above boundary conditions

$$V(x, y) = \sum_{i=0}^{\infty} \frac{\cos[u_i(1-x/d)]}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \quad (2.2.21)$$

$$\text{where } A_i = \frac{1}{d} \int_0^d V(x, 0) \cos[u_i(1-x/d)] dx \quad (2.2.22)$$

$$B_i = \frac{1}{d} \int_0^d V(x, L) \cos[u_i(1-x/d)] dx \quad (2.2.23)$$

$$\alpha_i = \frac{1}{2} \left[1 + \frac{\sin(2 u_i)}{2 u_i} \right] \quad (2.2.24)$$

and where the eigenvalues u_i are defined by

$$\tan u_i = R t_{ox} \frac{d}{u_i} \quad (2.2.25)$$

which arises from applying the top and bottom boundary conditions (2.2.13) and (2.2.17).

There is an infinite number of the eigenvalues u_i because $\tan u_i$ has an infinite number of branches and the hyperbola $d/(R t_{ox} u_i)$ intersects each branch once as illustrated in Figure 2.2.1. For typical device parameters the fundamental eigenvalue u_0 is approximately unity, and for large i the u_i are approximately $i\pi$ because the hyperbola $d/(R t_{ox} u_i)$ is nearly zero. To calculate the u_i exactly first write (2.2.25) as

$$\tan (u_i - i\pi) = \frac{a}{u_i} \quad (2.2.26)$$

$$\text{where } a = \frac{d}{R t_{ox}} \quad (2.2.27)$$

to recognize that $\tan u_i$ has an infinite number of branches. An initial guess for the u_i is obtained by approximating $\tan (u_i - i\pi)$ as the family of straight lines

$$\tan (u_i - i\pi) \approx u_i - i\pi \quad (2.2.28)$$

This is a good approximation because $\tan (u_i - i\pi)$ is roughly linear near the u_i axis where $d/(R t_{ox} u_i)$ intersects it.

2

Micro-D

Figure 2.2.1. Intersection points of $y = \tan u_i$ and the hyperbola $y = d/(R t_{ox} u_i)$.

Substitute (2.2.28) into (2.2.26), solve the resulting quadratic equation, and the following initial guess for the u_i is obtained:

$$u_i = \frac{i''}{2} + \sqrt{\left(\frac{i''}{2}\right)^2 + a} \quad (2.2.29)$$

Rearranging (2.2.26) to

$$u_i - i'' - \tan^{-1}(a/u_i) = 0 \quad (2.2.30)$$

and differentiating this expression produces the following Newton-Raphson iterative equation

$$u_i' = u_i - \frac{u_i - i'' - \tan^{-1}(a/u_i)}{1 + \frac{1}{u_i^2/a+a}} \quad (2.2.31)$$

where u_i' is the next estimate of the correct value of u_i . The u_i can be calculated to 7 digit accuracy by using (2.2.29) as an initial guess and iterating using (2.2.31) one or two times, although three iterations may be necessary to calculate u_0 .

The equations for A_i and B_i (2.2.22) and (2.2.23) result from evaluating (2.2.21) at $y = 0$ and $y = L$ respectively, i.e.

$$V(x, 0) = \sum_{i=0}^{\infty} \frac{\cos [u_i (1 - x/d)]}{\alpha_i} A_i \quad (2.2.32)$$

$$V(x, L) = \sum_{i=0}^{\infty} \frac{\cos [\bar{u}_i (1 - x/d)]}{\alpha_i} B_i \quad (2.2.33)$$

Multiplying each side of these two equations by $\cos [u_k (1-x/d)]$, integrating from 0 to d, and interchanging the order of summation and integration yields

$$\int_0^d V(x,0) \cos [u_k (1-x/d)] dx = \sum_{i=0}^{\infty} \frac{A_i}{\alpha_i} \int_0^d \cos [u_i (1-x/d)] \cos [u_k (1-x/d)] dx \quad (2.2.34)$$

$$\int_0^d V(x,L) \cos [u_k (1-x/d)] dx = \sum_{i=0}^{\infty} \frac{B_i}{\alpha_i} \int_0^d \cos [u_i (1-x/d)] \cos [u_k (1-x/d)] dx \quad (2.2.35)$$

The orthogonality property of the eigenfunctions $\cos [u_i (1-x/d)]$, namely

$$\int_0^d \cos [u_i (1-x/d)] \cos [u_k (1-x/d)] dx = d \alpha_i \delta_{ik} \quad (2.2.36)$$

where α_i is given by (2.2.24) and δ_{ik} is a Kronecker delta, is then taken advantage of to eliminate the summation signs. Upon rearranging (2.2.22) and (2.2.23) result.

Using the side boundary conditions specified by (2.2.18) and (2.2.19) in (2.2.22) and (2.2.23), closed form expressions for the coefficients A_i and B_i are obtained. This is because the side boundary conditions for $\psi(x,y)$ (2.1.39) and (2.1.40) have the trial function $T(x)$ as a simple additive term, and within it are contained all the nonpolynomial functions such as Gaussians and error functions. When (2.2.18) and (2.2.19) are derived from (2.1.39) and (2.1.40) using (2.2.16) $T(x)$ cancels taking with it all the nonpolynomial functions. Hence the expressions for

$V(x,0)$ and $V(x,L)$ are always cubic polynomials regardless of the substrate doping profile as can be observed from (2.2.18) and (2.2.19). The advantage of having polynomial expressions for $V(x,0)$ and $V(x,L)$ is that when they are substituted in (2.2.22) and (2.2.23) the integrals involving x^n and $\cos [u_i (1-x/d)]$ have closed form answers as shown below

$$\frac{1}{d} \int_0^d \cos [u_i (1 - x/d)] dx = \frac{\sin u_i}{u_i} \quad (2.2.37)$$

$$\frac{1}{d} \int_0^d x \cos [u_i (1 - x/d)] dx = d \frac{1 - \cos u_i}{u_i^2} \quad (2.2.38)$$

$$\frac{1}{d} \int_0^d x^2 \cos [u_i (1 - x/d)] dx = 2d^2 \frac{u_i - \sin u_i}{u_i^3} \quad (2.2.39)$$

$$\frac{1}{d} \int_0^d x^3 \cos [u_i (1 - x/d)] dx = 3d^3 \frac{u_i^2 + 2 (\cos u_i - 1)}{u_i^4} \quad (2.2.40)$$

Thus when (2.2.18) and (2.2.19) are substituted into (2.2.22) and (2.2.23), and the integrals performed using (2.2.37) to (2.2.40), the following closed form expressions for A_i and B_i are obtained

$$\begin{aligned}
A_i = & (V_{MS} - V_{GB} - R t_{ox} [F(0) - F(d) - Kd]) \frac{\sin u_i}{u_i} \\
& + \frac{d}{u_i^2} [(F_S - \frac{2V_0}{x_S} + Kd)(1 - \cos u_i) + (\frac{V_0}{x_S^2} - \frac{2F_S}{x_S} - \frac{K}{2})2d(1 - \frac{\sin u_i}{u_i}) \\
& + \frac{F_S}{x_S^2} 3d^2 (1 + 2 \frac{\cos u_i - 1}{u_i^2})] \quad (2.2.41)
\end{aligned}$$

$$\begin{aligned}
B_i = & (V_{MD} - V_{GB} - R t_{ox} [F(0) - F(d) - Kd]) \frac{\sin u_i}{u_i} \\
& + \frac{d}{u_i^2} [(F_D - \frac{2V_L}{x_D} + Kd)(1 - \cos u_i) + (\frac{V_L}{x_D^2} - \frac{2F_D}{x_D} - \frac{K}{2})2d(1 - \frac{\sin u_i}{u_i}) \\
& + \frac{F_D}{x_D^2} 3d^2 (1 + 2 \frac{\cos u_i - 1}{u_i^2})] \quad (2.2.42)
\end{aligned}$$

Since the quadratic and other simpler side boundary conditions are just special cases of the cubic boundary conditions, their A_i and B_i coefficients can be obtained by making simplifications to (2.2.41) and (2.2.42). For example, A_i and B_i for the quadratic boundary conditions specified by (2.1.17) and (2.1.18) are given by setting F_S and F_D to zero. Thus:

$$\begin{aligned}
 A_i = & (V_{MS} - V_{GB} - R t_{ox} [F(0) - F(d) - Kd]) \frac{\sin u_i}{u_i} \\
 & + \frac{d}{u_i^2} \left[(Kd - \frac{2V_0}{x_S}) (1 - \cos u_i) + (\frac{V_0}{2} - \frac{K}{2}) 2d (1 - \frac{\sin u_i}{u_i}) \right]
 \end{aligned}
 \quad (2.2.43)$$

$$\begin{aligned}
 B_i = & (V_{MD} - V_{GB} - R t_{ox} [F(0) - F(d) - Kd]) \frac{\sin u_i}{u_i} \\
 & + \frac{d}{u_i^2} \left[(Kd - \frac{2V_L}{x_D}) (1 - \cos u_i) + (\frac{V_L}{2} - \frac{K}{2}) 2d (1 - \frac{\sin u_i}{u_i}) \right]
 \end{aligned}
 \quad (2.2.44)$$

Coefficients that are comparable to those derived by Ratnakumar and Meindl [12], and Poole and Kwong [39,40], who assumed infinitely deep junctions, but are still valid for a nonuniformly doped substrate may be obtained by setting x_S and x_D to infinity and V_C to zero. This gives

$$A_i = (V_S - V_{GB} - R t_{ox} [F(0) - F(d)]) + \frac{Kd^2}{u_i^2} \frac{\sin u_i}{u_i} \quad (2.2.45)$$

$$B_i = (V_D - V_{GB} - R t_{ox} [F(0) - F(d)]) + \frac{Kd^2}{u_i^2} \frac{\sin u_i}{u_i} \quad (2.2.46)$$

Finally, if the substrate is uniformly doped then $F(0) = F(d)$ and

$$A_i = (V_S - V_{GB} + \frac{Kd^2}{u_i^2}) \frac{\sin u_i}{u_i} \quad (2.2.47)$$

$$B_i = (V_D - V_{GB} + \frac{Kd^2}{u_i^2}) \frac{\sin u_i}{u_i} \quad (2.2.48)$$

These coefficients are identical to those derived by Poole and Kwong [39]. (Note that there is a typographical error in equation (6) of [39] for C_n . Once the differences in notation are allowed for (2.2.48) provides the correct form of equation (6)).

The solution for $V(x,y)$ has now been completely presented. All that remains to finish the solution for $\psi(x,y)$ is to sum $T(x)$, $U(x)$ and $V(x,y)$. Thus substituting (2.2.7), (2.2.15) and (2.2.21) into (2.1.5) $\psi(x,y)$ is obtained as

$$\begin{aligned} \psi(x,y) = & G(x) - G(d) + (d-x) F(d) + \frac{K}{2} (x^2 - 2dx) + V_{GB} \\ & - T(0) + R t_{ox} [F(0) - F(d) - Kd] \\ & + \sum_{i=0}^{\infty} \frac{\cos [u_i (1-x/d)]}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \end{aligned} \quad (2.2.49)$$

This equation is the solution of the Poisson problem for a short-channel MOSFET in subthreshold. From it can be obtained virtually all information about the behavior of the device in this regime.

To demonstrate the nature of this solution in a more vivid way than is possible with just equations, three-dimensional pictures of the two-dimensional potential are used. Plots of

$T(x)$, $U(x)$, $V(x,y)$ and $\psi(x,y)$ are shown for a short-channel device with a channel length of $1\text{ }\mu\text{m}$ in Figure 2.2.2, and for a long-channel device with a channel length of $10\text{ }\mu\text{m}$ in Figure 2.2.3. Both are CMOS devices whose device parameters are listed in Table 2.1.1, and they are identically biased at $V_{GS} = 0.8\text{ V}$, $V_{DS} = 5\text{ V}$ and $V_{SB} = 0$. The most interesting contrast between the two plots of $\psi(x,y)$ is that for the $10\text{ }\mu\text{m}$ device the long-channel solution $T(x) + U(x)$ appears clearly in the center of the device and the influence of $V(x,y)$ is confined to the edges near the source and drain. But for the $1\text{ }\mu\text{m}$ device the long-channel solution $T(x) + U(x)$ is overlaid by $V(x,y)$ everywhere. The reason for this is that $V(x,y)$ embodies the influence of the source and drain in pulling up the potential which is the cause of short-channel effects. For a long-channel device the source and drain are able to influence the potential only along the edges and hence have little net effect on device behavior. But for a short-channel device the source and drain pull up the potential everywhere in the channel and below, and hence have a major effect on device behavior. This is the essence of short-channel effects.

2.3 The Threshold Voltage Model

The two-dimensional potential $\psi(x,y)$ specified by (2.2.49) can be used to derive a short-channel threshold voltage model. This is done by isolating the V_{GB} dependence of $\psi(x,y)$ and

(a)

(b)

(c)

(d)

Figure 2.2.2. (a) $T(x)$, (b) $U(x)$, (c) $V(x, y)$ and (d) $\psi(x, y)$ for a CMOS device where $L = 1 \mu\text{m}$, $V_{GS} = 0.8 \text{ V}$, $V_{DS} = 5 \text{ V}$ and $V_{SB} = 0$.

(a)

(b)

(c)

(d)

Figure 2.2.3. (a) $T(x)$, (b) $U(x)$, (c) $V(x, y)$ and (d) $\psi(x, y)$ for a CMOS device where $L = 10 \text{ } \mu\text{m}$, $V_{GS} = 0.8 \text{ V}$, $V_{DS} = 5 \text{ V}$, $V_{SB} = 0$.

setting the minimum surface potential to $2\phi_p + V_{SB}$ which is its value at threshold. The $2\phi_p + V_{SB}$ condition is a widely used definition of threshold but its use for short-channel devices is somewhat suspect because it is a long-channel condition. Since there is no alternative this condition is used anyway and its shortcomings are discussed in Chapter 3. To isolate the V_{GB} dependence of $\psi(x,y)$ let

$$A_i = C_i - V_{GB} E_i \quad (2.3.1)$$

$$B_i = D_i - V_{GB} F_i \quad (2.3.2)$$

and hence

$$\begin{aligned} \psi(x,y) = & G(x) - G(d) + (d-x) F(d) + \frac{K}{2} (x^2 - 2dx) - T(0) \\ & + R t_{ox} [F(0) - F(d) - Kd] \\ & + \sum_{i=0}^{\infty} \frac{\cos [u_i (1-x/d)]}{\alpha_i \sinh (u_i L/d)} [C_i \sinh (u_i \frac{L-y}{d}) + D_i \sinh (u_i y/d)] \\ & + V_{GB} (1 - \sum_{i=0}^{\infty} \frac{\cos [u_i (1-x/d)]}{\alpha_i \sinh (u_i L/d)} [E_i \sinh (u_i \frac{L-y}{d}) + F_i \sinh (u_i y/d)]) \end{aligned} \quad (2.3.3)$$

where

$$\begin{aligned}
 C_i = & (V_{MS} - R t_{ox} [F(0) - F(d) - Kd]) \frac{\sin u_i}{u_i} \\
 & + \frac{d}{u_i^2} \left[(G_S - \frac{2V_0}{x_S} + Kd)(1 - \cos u_i) + (\frac{V_0}{x_S^2} - \frac{2G_S}{x_S} - \frac{K}{2}) 2d(1 - \frac{\sin u_i}{u_i}) \right. \\
 & \left. + \frac{G_S}{x_S^2} 3d^2 (1 + \frac{2(\cos u_i - 1)}{u_i^2}) \right] \quad (2.3.4)
 \end{aligned}$$

$$\begin{aligned}
 D_i = & (V_{MD} - R t_{ox} [F(0) - F(d) - Kd]) \frac{\sin u_i}{u_i} \\
 & + \frac{d}{u_i^2} \left[(G_D - \frac{2V_L}{x_D} + Kd)(1 - \cos u_i) + (\frac{V_L}{x_D^2} - \frac{2G_D}{x_D} - \frac{K}{2}) 2d(1 - \frac{\sin u_i}{u_i}) \right. \\
 & \left. + \frac{G_D}{x_D^2} 3d^2 (1 + \frac{2(\cos u_i - 1)}{u_i^2}) \right] \quad (2.3.5)
 \end{aligned}$$

$$\begin{aligned}
 E_i = & \frac{\sin u_i}{u_i} + \frac{\delta_{S1} \delta_{S2} d}{R t_{ox} u_i^2} [1 - \cos u_i - \frac{4d}{x_S} (1 - \frac{\sin u_i}{u_i}) \\
 & + \frac{3d^2}{x_S^2} (1 + \frac{2(\cos u_i - 1)}{u_i^2})] \quad (2.3.6)
 \end{aligned}$$

$$F_i = \frac{\sin u_i}{u_i} + \frac{\delta_{D1} \delta_{D2} d}{R t_{ox} u_i^2} \left[1 - \cos u_i - \frac{4d}{x_D} \left(1 - \frac{\sin u_i}{u_i} \right) + \frac{3d^2}{x_D^2} \left(1 + \frac{2(\cos u_i - 1)}{u_i^2} \right) \right] \quad (2.3.7)$$

$$G_S = \left[\frac{V_{MS}}{R t_{ox}} \delta_{S1} + \frac{2V_0}{x_S} + F(d) - F(0) \right] \delta_{S2} \quad (2.3.8)$$

$$G_D = \left[\frac{V_{MD}}{R t_{ox}} \delta_{D1} + \frac{2V_L}{x_D} + F(d) - F(0) \right] \delta_{D2} \quad (2.3.9)$$

where these expressions have extracted the V_{GB} dependence of A_i and B_i from equations (2.2.41) and 2.2.42). Note that both the explicit V_{GB} dependence and the V_{GB} dependence of F_S and F_D have been isolated. Equations (2.3.4) to (2.3.9) can easily be reduced to their forms for the simpler side boundary conditions as demonstrated for A_i and B_i in the last section. For the quadratic boundary conditions this is done by setting δ_{S2} and δ_{D2} to zero, which is the equivalent of setting F_S and F_D to zero. If infinitely deep junctions are assumed then the simplifications that are necessary are to set x_S and x_D to infinity and V_C to zero. Finally, if a uniform substrate is assumed then $F(d)$ is set equal to $F(0)$ and $T(0)$ is set equal to zero.

To derive an expression for the threshold voltage the following three conditions must be applied.

$$\psi(0, y_m) = 2\phi_p + V_{SB} \quad (2.3.10)$$

$$\psi(d, y_m) = 0 \quad (2.3.11)$$

$$\frac{\partial \psi}{\partial y}(0, y_m) = 0 \quad (2.3.12)$$

$$\text{where } \phi_p = \frac{kT}{q} \ln \left[\frac{N_B + N_n(d)}{n_i} \right] \quad (2.3.13)$$

and where y_m is the y coordinate of the minimum surface potential. The first condition (2.3.10) is the long-channel condition for threshold. As mentioned previously, it is certainly debatable whether it applies to short-channel devices and this is discussed in Chapter 3, but it is used here for lack of a better alternative. The second condition (2.3.11) specifies a zero potential at the minimum depletion depth point (d, y_m) (see Figure 1.2.1). This is the remaining depletion edge condition as a zero field at (d, y_m) has already been specified by (2.1.7), and (2.3.11) is used to determine the minimum depletion depth d . The third condition (2.3.12) states that $(0, y_m)$ is the position of the minimum potential on the surface. This assumes that the minimum surface potential point $(0, y_m)$ is directly above the minimum depletion depth point (d, y_m) , which in general is not exactly true but is a good approximation.

Applying (2.3.10) and (2.3.11) to (2.3.3), and applying (2.3.12) to (2.2.49) yields

$$\psi(0, y_m) = 2\phi_p + V_{SB} = R t_{ox} [F(0) - F(d) - Kd]$$

$$+ \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [C_i \sinh(u_i \frac{L-y_m}{d}) + D_i \sinh(u_i y_m/d)]$$

$$+ V_{GB} (1 - \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [E_i \sinh(u_i \frac{L-y_m}{d}) + F_i \sinh(u_i y_m/d)])$$

(2.3.14)

$$\psi(d, y_m) = 0 = R t_{ox} [F(0) - F(d) - Kd] - \frac{K}{2} d^2 - T(0)$$

$$+ \sum_{i=0}^{\infty} \frac{1}{\alpha_i \sinh(u_i L/d)} [C_i \sinh(u_i \frac{L-y_m}{d}) + D_i \sinh(u_i y_m/d)]$$

$$+ V_{GB} (1 - \sum_{i=0}^{\infty} \frac{1}{\alpha_i \sinh(u_i L/d)} [E_i \sinh(u_i \frac{L-y_m}{d}) + F_i \sinh(u_i y_m/d)])$$

(2.3.15)

$$\frac{\partial \psi}{\partial y}(0, y_m) = 0 = g(y_m) = \frac{1}{d} \sum_{i=0}^{\infty} \frac{u_i \cos u_i}{\alpha_i \sinh(u_i L/d)} [B_i \cosh(u_i y_m/d) - A_i \cosh(u_i \frac{L-y_m}{d})]$$

(2.3.16)

This is a system of three equations in the three unknowns V_{GB} , d and y_m . The equations are linear in V_{GB} and nonlinear in d and y_m . Therefore a closed form expression for V_{GB} may be obtained but Newton-Raphson iterative equations must be derived to solve for d and y_m . The Newton-Raphson expressions are simultaneously used to solve for d and y_m , and once these quantities are found V_{GB} is calculated from its closed form equation. To form the

Newton-Raphson expression for y_m (2.3.16) is differentiated with respect to y_m giving

$$\frac{\partial g}{\partial y_m} = \frac{1}{d^2} \sum_{i=0}^{\infty} \frac{u_i^2 \cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y_m}{d}) + B_i \sinh(u_i y_m/d)] \quad (2.3.17)$$

Before proceeding further it is useful to simplify these equations by defining the following six summations

$$\Sigma_1 = \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [E_i \sinh(u_i \frac{L-y_m}{d}) + F_i \sinh(u_i y_m/d)] \quad (2.3.18)$$

$$\Sigma_2 = \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [C_i \sinh(u_i \frac{L-y_m}{d}) + D_i \sinh(u_i y_m/d)] \quad (2.3.19)$$

$$\Sigma_3 = \sum_{i=0}^{\infty} \frac{E_i \sinh(u_i \frac{L-y_m}{d}) + F_i \sinh(u_i y_m/d)}{\alpha_i \sinh(u_i L/d)} \quad (2.3.20)$$

$$\Sigma_4 = \sum_{i=0}^{\infty} \frac{C_i \sinh(u_i \frac{L-y_m}{d}) + D_i \sinh(u_i y_m/d)}{\alpha_i \sinh(u_i L/d)} \quad (2.3.21)$$

$$\Sigma_5 = \sum_{i=0}^{\infty} \frac{u_i \cos u_i}{\alpha_i \sinh(u_i L/d)} [B_i \cosh(u_i y_m/d) - A_i \cosh(u_i \frac{L-y_m}{d})] \quad (2.3.22)$$

$$\Sigma_6 = \sum_{i=0}^{\infty} \frac{u_i^2 \cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y_m}{d}) + B_i \sinh(u_i y_m/d)] \quad (2.3.23)$$

Equations (2.3.14) to (2.3.17) then become

$$2\phi_p + V_{SB} = R t_{ox} [F(0) - F(d) - Kd] + \Sigma_2 + V_{GB} (1 - \Sigma_1) \quad (2.3.24)$$

$$0 = R t_{ox} [F(0) - F(d) - Kd] - \frac{K}{2} d^2 - T(0) + \Sigma_4 + V_{GB} (1 - \Sigma_3) \quad (2.3.25)$$

$$g(y_m) = \Sigma_5/d \quad (2.3.26)$$

$$\frac{\partial g}{\partial y_m} = \Sigma_6/d^2 \quad (2.3.27)$$

The closed form equation for V_{GB} is obtained by rearranging (2.3.24) to,

$$V_{GB} = \frac{2\phi_p + V_{SB} + R t_{ox} [F(d) - F(0) + Kd] - \Sigma_2}{1 - \Sigma_1} \quad (2.3.28)$$

Since at threshold $V_{GS} = V_T$ then

$$V_{GB} = V_T - V_{FB} + V_{SB} \quad (2.3.29)$$

and thus the closed form expression for the threshold voltage is found to be

$$V_T = V_{FB} - V_{SB} + \frac{2\phi_p + V_{SB} + R t_{ox} [F(d) - F(0) + Kd] - \Sigma_2}{1 - \Sigma_1} \quad (2.3.30)$$

Substituting (2.3.28) into (2.3.25) gives

$$\begin{aligned}
 0 = h(d) = & R t_{ox} [F(0) - F(d) - Kd] - \frac{K}{2} d^2 - T(0) + \Sigma_4 \\
 & + \frac{1 - \Sigma_3}{1 - \Sigma_1} (2\phi_p + V_{SB} + R t_{ox} [F(d) - F(0) + Kd] - \Sigma_2)
 \end{aligned}
 \tag{2.3.31}$$

Neglecting the d dependence of Σ_1 , Σ_2 , Σ_3 and Σ_4 and differentiating (2.3.31) with respect to d , yields the following approximate derivative that is needed for the Newton-Raphson expression in d

$$\frac{\partial h}{\partial d} = -[K + f(d)] \left(d + R t_{ox} \frac{\Sigma_3 - \Sigma_1}{1 - \Sigma_1} \right)
 \tag{2.3.32}$$

Using an approximate derivative such as (2.3.32) generally results in slower convergence of the iterative process. But in this case the extra calculations needed for the additional iterations are more than offset by the calculations saved by using a simpler derivative. The overall effect is that the iterative process to find d is faster when the approximate derivative given by (2.3.32) is used. Combining (2.3.31) and (2.3.32) gives the Newton-Raphson expression that is used to calculate d

$$d' = d - \left(\frac{K}{2} d^2 + T(0) - \Sigma_4 - \frac{1 - \Sigma_3}{1 - \Sigma_1} (2\phi_p + V_{SB} - \Sigma_2) \right. \\ \left. + R t_{ox} [F(d) - F(0) + Kd] \frac{\Sigma_3 - \Sigma_1}{1 - \Sigma_1} \right) / \{ [K + f(d)] (d + R t_{ox} \frac{\Sigma_3 - \Sigma_1}{1 - \Sigma_1}) \} \quad (2.3.33)$$

Similarly, combining (2.3.26) and (2.3.27) gives the Newton-Raphson expression that is used to calculate y_m .

$$y_m' = y_m - d \frac{\Sigma_5}{\Sigma_6} \quad (2.3.34)$$

Since (2.3.33) and (2.3.34) are coupled transcendental expressions they are solved by starting with the initial guesses

$$d = \sqrt{\frac{2\phi_p + V_{SB}}{M}} \quad (2.3.35)$$

$$y_m = \frac{L}{2} \quad (2.3.36)$$

then using each alternately until both d and y_m converge. Once d and y_m are known then the threshold voltage can be calculated from (2.3.30). Between 1 and 25 Newton-Raphson iterations are required to calculate d and y_m , with the number of iterations increasing as short-channel effects become more severe. Convergence is always obtained unless the depletion depth d is driven below the depth x_g by the severity of the short-channel effects. Below x_g the source side boundary condition ceases to be valid and the calculation of the threshold voltage is

abandoned. The CPU time requirements to calculate the threshold voltage are between 1 and 45 milliseconds on an IBM 3081KX computer where the computation time increases with the number of iterations required. The validity and accuracy of this threshold voltage model is demonstrated by comparison to the results of MINIMOS simulations in Chapter 3.

2.4 The Subthreshold Current Model

The two-dimensional potential solution described by (2.2.49) may also be used as the basis of a short-channel subthreshold drain current model. The first analytical model of this type was proposed by Ratnakumar and Meindl [12]. This was a very simplistic model that did nothing more than translate the long-channel subthreshold characteristic along the V_{GS} axis by the short-channel threshold voltage reduction ΔV_T . A more sophisticated model was published by Poole and Kwong [40] who calculated the drain current I_D by integrating the surface electron density from the source to the drain. The resulting drain current expression is simple but it makes the assumption of a finite thickness for the inversion layer, which is calculated from the normal field at the surface. An alternate method to calculate the subthreshold current, where the only major assumption is that the current flow is one-dimensional, is to use the equation derived by De La Moneda [8] from the current continuity equation, that is based on Gummel's algorithm [30]. This approach is used here.

The derivation of the drain current expression starts with the equation for the mobile charge density per unit area of the channel Q_m :

$$Q_m = -q n_i \int_0^\infty [e^{\beta(\psi - \phi_n)} - e^{-\beta\phi_n}] dx \quad (2.4.1)$$

It is supposed that Q_m is sufficiently small as to be negligible in the solution of Poisson's equation, and therefore ψ is the potential given by (2.2.49) that was derived on the basis of complete depletion. Assuming that the current flow is one-dimensional in the y direction makes ϕ_n independent of x , and hence the term $e^{-\beta\phi_n}$ may be factored out of the integral giving

$$Q_m = -q n_i e^{-\beta\phi_n} \int_0^\infty (e^{\beta\psi} - 1) dx \quad (2.4.2)$$

$$\text{and letting } z = e^{-\beta\phi_n} \quad (2.4.3)$$

$$f = \int_0^\infty (e^{\beta\psi} - 1) dx \quad (2.4.4)$$

allows (2.4.2) to be written as

$$Q_m = -q n_i z f \quad (2.4.5)$$

Differentiating (2.4.3) gives

$$\frac{d\phi_n}{dy} = -\frac{1}{\beta z} \frac{dz}{dy} \quad (2.4.6)$$

With the assumption of one-dimensional current flow in the y direction and that the mobility is independent of x, the drain current is given by

$$I_D(y) = - \mu_n W Q_m \frac{d\phi_n}{dy} \quad (2.4.7)$$

where μ_n is the mobility and W is the width of the MOSFET. Substituting (2.4.5) and (2.4.6) into (2.4.7) yields

$$I_D(y) = - \frac{W q n_i}{B} \mu_n f \frac{dz}{dy} \quad (2.4.8)$$

Current continuity is invoked by setting the product of the terms of (2.4.8) that depend upon y equal to a constant A.

$$\mu_n f \frac{dz}{dy} = A \quad (2.4.9)$$

Rearranging and integrating along the channel gives

$$z(y) = z(0) + A \int_0^y \frac{dy}{\mu_n f} \quad (2.4.10)$$

This expression is now evaluated at $y = L$ and substituting the values of $\phi_n(0)$ and $\phi_n(L)$ which are

$$\phi_n(0) = \phi_p + V_{SB} \quad (2.4.11)$$

$$\phi_n(L) = \phi_p + V_{SB} + V_{DS} \quad (2.4.12)$$

yields

$$A = \frac{(e^{-\beta V_{DS}} - 1) e^{-\beta(\phi_p + V_{SB})}}{\int_0^L \frac{dy}{u_n f}} \quad (2.4.13)$$

After substituting (2.4.13) into (2.4.9), and (2.4.9) into (2.4.8), the following expression for the drain current is obtained

$$I_D = \frac{W q n_i}{\beta} (1 - e^{-\beta V_{DS}}) e^{-\beta(\phi_p + V_{SB})} \int_0^L \frac{dy}{u_n \int_0^\infty (e^{\beta\psi} - 1) dx} \quad (2.4.14)$$

The major assumption in the derivation of (2.4.14) is that the current flow is one-dimensional. This is assumed in (2.4.2) because $e^{-\beta\phi_n}$ is removed from the integral, and in (2.4.7) by making the gradient of the quasi-Fermi potential ϕ_n only in the y direction. The assumption of one-dimensional current flow is violated whenever enough drain voltage is applied to raise the potential of the part of the channel near the drain above the gate potential. This is almost always the case in subthreshold because the gate potential is low. When the potential in this part of the channel rises above the gate potential then the normal field at the surface reverses in this region and drives the electrons below the surface, thereby producing two-dimensional current flow. So the assumption of

one-dimensional current flow is only valid from the source to the point at which field reversal begins called the field reversal point whose y coordinate is L' . It is thus only valid to integrate (2.4.9) from 0 to L' , but if this is done then $\phi_n(L')$ is not known. It turns out that this does not matter because if V_{DS} is large then

$$e^{-\beta\phi_n(L')} \ll e^{-\beta\phi_n(0)} \quad (2.4.15)$$

in which case the value of $\phi_n(L')$ is of no consequence, and if V_{DS} is small then

$$L \approx L' \quad (2.4.16)$$

$$\text{and } \phi_n(L') \approx \phi_n(L) \quad (2.4.17)$$

Thus to a very good approximation L can be switched to L' in (2.4.14) giving

$$I_D = \frac{W q n_i}{\beta} \frac{(1 - e^{-\beta V_{DS}}) e^{-\beta(\phi_p + V_{SB})}}{\int_0^{L'} \frac{dy}{u_n \int_0^\infty (e^{\beta\psi} - 1) dx}} \quad (2.4.18)$$

To calculate the position of the field reversal point L' (2.2.49) is evaluated at $(0, L')$ and equated to V_{GB}' .

$$\psi(0, L') = V_{GB}' = V_{GB}' - R t_{ox} [F(d) - F(0) - Kd] + \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-L'}{d}) + B_i \sinh(u_i L'/d)] \quad (2.4.19)$$

Defining

$$p(L') = R t_{ox} [F(d) - F(0) + Kd] - \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-L'}{d}) + B_i \sinh(u_i L'/d)] \quad (2.4.20)$$

permits L' to be calculated by a Newton-Raphson solution of $p(L') = 0$. Differentiating (2.4.20)

$$\frac{\partial p}{\partial L'} = - \frac{1}{d} \sum_{i=0}^{\infty} \frac{u_i \cos u_i}{\alpha_i \sinh(u_i L/d)} [B_i \cosh(u_i L'/d) - A_i \cosh(u_i \frac{L-L'}{d})] \quad (2.4.21)$$

and defining the sums

$$\Sigma_8 = \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-L'}{d}) + B_i \sinh(u_i L'/d)] \quad (2.4.22)$$

$$\Sigma_9 = \sum_{i=0}^{\infty} \frac{u_i \cos u_i}{\alpha_i \sinh(u_i L/d)} [B_i \cosh(u_i L'/d) - A_i \cosh(u_i \frac{L-L'}{d})] \quad (2.4.23)$$

allows (2.4.20) and (2.4.21) to be combined to produce the following Newton-Raphson equation for the calculation of L' .

$$L'' = L' + d \frac{R t_{ox} [F(d) - F(0) + Kd] - \Sigma_8}{\Sigma_9} \quad (2.4.24)$$

where L'' is the next estimate of the true value of L' in the iterative sequence. The initial guess to be used with (2.4.24) comes from an adaptation of the Reddi and Sah [21] model and is expressed as

$$L' = L - 0.55 \sqrt{\frac{V_{MD}}{M} - V_{GB}} \quad (2.4.25)$$

where V_{MD} and M are given by (2.1.12) and (2.1.46) respectively.

Just as with the threshold voltage model a quantity that must be calculated to use the subthreshold current model is the minimum depletion depth d . The situation here is simpler than previously because V_{GB} is known and hence only the two conditions

$$\psi(d, y_m) = 0 \quad (2.4.26)$$

$$\frac{\partial \psi}{\partial y}(0, y_m) = 0 \quad (2.4.27)$$

need be considered to calculate the two unknowns d and y_m . Using (2.4.26) with (2.2.49)

$$\begin{aligned} \psi(d, y_m) = 0 = R t_{ox} [F(0) - F(d) - Kd] - \frac{K}{2} d^2 - T(0) + V_{GB}' \\ + \sum_{i=0}^{\infty} \frac{A_i \sinh(u_i \frac{L-y_m}{d}) + B_i \sinh(u_i y_m/d)}{\alpha_i \sinh(u_i L/d)} \end{aligned} \quad (2.4.28)$$

$$\text{Defining } \Sigma_7 = \sum_{i=0}^{\infty} \frac{A_i \sinh(u_i \frac{L-y_m}{d}) + B_i \sinh(u_i y_m/d)}{\alpha_i \sinh(u_i L/d)} \quad (2.4.29)$$

and differentiating (2.4.28) permits the construction of the following Newton-Raphson expression.

$$d' = d + \frac{R t_{ox} [F(0) - F(d) - Kd] - K/2 d^2 - T(0) + V_{GB}' + \Sigma_7}{[K + f(d)] (d + R t_{ox})} \quad (2.4.30)$$

The procedure for determining d and y_m is much the same as for the threshold voltage model. Initial guesses are made using (2.3.35) and (2.3.36) and then (2.3.34) and (2.4.30) are used alternately until the values of d and y_m converge. Once d and y_m are determined, L' is calculated from (2.4.24) and (2.4.25). The sequence of calculations concludes with the calculation of I_D using (2.4.18). Note that (2.4.18) requires the calculation of a double numerical integral. This can be done quite efficiently using Gaussian quadrature and it has been found that a 10 point formula provides sufficient accuracy. Thus only 100 evaluations of $\psi(x, y)$ are necessary to perform the double integral and calculate the drain current. The validity and accuracy of the subthreshold model is checked in Chapter 3 by the comparison of

simulated subthreshold characteristics with MINIMOS simulations
and measurements.

CHAPTER 3

VERIFICATION OF THE SUBTHRESHOLD MODEL

3.1 Testing the Threshold Voltage Model

The validity and accuracy of an analytical device model may be tested by comparing it to measurements or the results of numerical simulations. It is certainly preferable to use measurements since the ultimate objective of device modelling is to simulate the real world, but there are some cases where it is necessary to use numerical simulations instead. This is admissible if it is known that the numerical model accurately reflects the behavior of the real device, which is well established for the MINIMOS model. For the verification of a short-channel threshold voltage model it is necessary to use a numerical model for two reasons.

The primary reason is that the threshold condition

$$\psi(0, y_m) = 2\phi_p + V_{SB} \quad (3.1.1)$$

is well related to measured threshold voltages only for long-channel devices. For short-channel devices Toyabe and Asai [36] found from numerical simulations that the factor of 2 must be reduced to a number between 1 and 2, and this number can only be related to the device parameters through the use of fitting parameters. This fact has been generally ignored in analytical short-channel threshold voltage models [10-12, 41-45] because there is no good alternative to (3.1.1). This would be less of a

problem if it were possible to probe inside a MOSFET and measure its surface potential but unfortunately this cannot be done. The second reason for using a numerical model is that identical device structures can then be simulated by the analytical and numerical models. If in addition (3.1.1) is used as the threshold condition by both models then the resulting threshold voltages are exactly comparable.

It is a valid question to ask if all analytical short-channel threshold voltage models use (3.1.1) to define threshold, but this condition is not related to measured threshold voltage models, then what is the use of such models? The answer to this question is that if (3.1.1) is used the resulting threshold voltages will still approximate the measured values although the error will become larger as short-channel effects increase. Furthermore, it is possible for short-channel threshold voltage models to define quite accurately the demarcation between the long and short-channel regimes. Yet another reason for their use is that short-channel threshold voltage models can be used to include charge sharing in modified long-channel models used for circuit simulation. But their major use in this thesis is as a method to check the accuracy of the analytical solution for the two-dimensional potential $\psi(x,y)$. The task of checking the accuracy of $\psi(x,y)$ against measurements can then be left to the subthreshold current model. In fact, it is the opinion of this author that the role of threshold voltages should be de-emphasized for short-channel devices because of the dubious validity of (3.1.1), and the role of the threshold

voltage should be replaced by the subthreshold current characteristic because it has a much stronger physical basis for short-channel devices.

To test the accuracy of the short-channel threshold voltage model MINIMOS simulations were performed using a variety of channel lengths for the uniform substrate, NMOS and CMOS devices whose device parameters are given in Table 2.1.1. In order to define threshold using (3.1.1) the built-in MINIMOS threshold voltage model, which uses a current level definition of threshold, was dispensed with. Instead multiple simulations were performed for each threshold voltage where the gate voltage was modified until condition (3.1.1) was satisfied. This is the method recommended by Viswanathan et al. [28] to obtain a threshold voltage from a numerical model. The resulting threshold voltage versus channel length curves for $V_{SB} = 0$ and 5 V were compared to the threshold voltages predicted by the analytical model when it uses cubic, quadratic and infinitely deep junction side boundary conditions. The results of this comparison are shown in Figures 3.1.1, 3.1.2 and 3.1.3 for the uniform substrate, NMOS and CMOS devices respectively. All of the threshold voltage versus channel length curves have been cut off at the channel length where the minimum depletion depth d exceeds the depletion depth x_s . This is because if d is greater than x_s then the cubic and quadratic boundary conditions cease to be valid.

It is clear from these figures that the cubic boundary condition curves are the most accurate, the quadratic boundary

Figure 3.1.1. Threshold voltage versus channel length for a uniform substrate device where $V_{DS} = 0.1$ V.

Figure 3.1.2. Threshold voltage versus channel length for an NMOS device where $V_{DS} = 0.1\text{ V}$.

Figure 3.1.3. Threshold voltage versus channel length for a CMOS device where $V_{DS} = 0.1$ V.

condition curves are less so, and the infinitely deep junction curves are the least inaccurate. The reason for this is that the accuracy of the analytical threshold voltage model is directly related to the accuracy of the side boundary conditions because they determine the amount of electric displacement entering the rectangular box from the source and drain, which is what is responsible for short-channel effects. The influence of the side boundary conditions is best explained by noting that Poisson's equation governs both the determination of a two-dimensional electrostatic potential and the deflection of a two-dimensional elastic membrane. Thus the $\psi(x,y)$ curves presented in Figures 2.2.2 and 2.2.3 could be thought of as the shape of an elastic membrane. If one side of a rectangular elastic membrane is pulled up while the conditions on the other three sides are kept the same, then it is intuitively obvious that the gradient along that side will increase. In electrostatic terms the gradient is the electric field and the gradient times the permittivity of the material is the electric displacement. So if the potential along the side of the box is increased, then the electric displacement crossing this boundary is also increased and vice versa. Hence if the approximate side boundary conditions are in general higher than the potential predicted by MINIMOS, then too much displacement will cross the side boundaries, the short-channel effect will be overestimated because too much of the substrate charge will be claimed by the source and drain, and the short-channel threshold voltage will

then be lower than that predicted by MINIMOS. The converse is also true because if the side boundary potentials are lower than those predicted by MINIMOS, then the displacement crossing the side boundaries will be underestimated and the short-channel threshold voltage will be too high. Using this argument it is clear that the reason the cubic boundary condition threshold voltage model is the most accurate is because the cubic boundary conditions give the best estimate of the potential variation on the sides of the box, and hence the best estimate of the displacement crossing these sides. As discussed in Section 2.1 the quadratic boundary conditions are less accurate than the cubic ones and hence the threshold voltages predicted using these boundary conditions are less accurate. The infinitely deep junction boundary conditions grossly overestimate the potential along, and the displacement crossing, the sides of the box, and hence they drastically overestimate the short-channel effect and underestimate the short-channel threshold voltage. This conclusion calls into question the claims of high accuracy made by Ratnakumar and Meindl [12] whose threshold voltage model assumed infinitely deep junctions.

The principles of linking the potential and electric displacement along the sides of the box to the magnitude of the threshold voltage can be applied in even greater detail. It is found that if any specific threshold voltage is too high compared to MINIMOS simulations then its side boundary conditions will be too low when compared to the same simulations, and vice versa. This is demonstrated by the cubic side boundary conditions $\psi(x,0)$

and $\psi(x,L)$ plotted in Figures 2.1.7, 2.1.8 and 2.1.9 which correspond to one particular threshold voltage on each of the six curves of Figures 3.1.1, 3.1.2 and 3.1.3. For the uniform substrate device with $L = 1.5 \mu\text{m}$ $V_{SB} = 5 \text{ V}$, NMOS device with $L = 1.5 \mu\text{m}$ $V_{SB} = 5 \text{ V}$, and CMOS device with $L = 0.8 \mu\text{m}$ $V_{SB} = 5 \text{ V}$ the side boundary conditions are in general too low when compared to the MINIMOS predictions, and as expected for these cases the threshold voltages on the cubic boundary condition curves are too high. The converse is true for the uniform substrate device with $L = 0.5 \mu\text{m}$ $V_{SB} = 0$ where the side boundary conditions are too high and the threshold voltage too low. For the NMOS device with $L = 1.2 \mu\text{m}$ $V_{SB} = 0$ the side boundary conditions are at some points too high and at others too low so the net error in the total displacement crossing the side boundary conditions is near zero and consequently the threshold voltage is in nearly exact agreement with MINIMOS. The most accurate case of the six is for the CMOS device with $L = 0.5 \mu\text{m}$ $V_{SB} = 0$ where both the side boundary conditions and threshold voltage are in near perfect agreement with MINIMOS.

Since it is possible to relate the error in the threshold voltage so closely to the error in the side boundary conditions, it is clear that the errors in the top and bottom boundary conditions are negligible compared to the error in the side boundary conditions. This is reasonable because the error for the bottom boundary condition is the deviation of the normal displacement from its assumed zero value, which is small since the bottom boundary is at or near the depletion edge. It is also

reasonable for the top boundary condition because its error is the neglected ~~mobile~~ charge, which in subthreshold is nonnegligible only in a small region very close to the source and drain where the surface potential is pulled up.

Thus it is concluded that if the cubic boundary conditions are used then all four boundary conditions are quite accurate. This results in an accurate threshold voltage model and solution for the two-dimensional potential. It is clear that if more accuracy is desired then it will most easily be gained by improving the side boundary conditions. This issue is addressed in Chapter 4.

3.2 Testing the Subthreshold Current Model

3.2a Determining the Device Parameters

Since it was deemed to be unreasonable to compare the subthreshold model to device measurements using threshold voltages, this task will be performed using subthreshold currents. For such a comparison to be meaningful it is essential that the devices to be measured are accurately characterized. This is in order to ensure that the accuracy of the model is tested and not its ability to use device parameters as fitting parameters. Even so, there is always an experimental error in the measured device parameters, and the best indication of where the true value lies within the range of possible values is in the fit of the measured and simulated results.

To fully characterize a MOSFET so that its subthreshold characteristics may be compared to those of the short-channel

subthreshold current model it is necessary to measure: the electrical channel length L , the device width W , the substrate doping profile $N(x)$, the peak doping in the source and drain N_S , the oxide thickness t_{ox} , the junction depth r_j , and the flatband voltage V_{FB} . The types of devices that were characterized are referred to here as Device Type 1 and Device Type 2. They are both n-channel devices and an array of devices of different channel lengths was available for each. An attempt was made to restrict the characterization procedures to electrical measurements performed on MOSFETs or MOS capacitors on the same die. This permitted most of the characterization to be done on the same devices from which the subthreshold current measurements were taken.

The electrical channel lengths were measured by Whitfield's method [51]. This method uses a pair of devices of different channel lengths, which are otherwise identical, and extracts the electrical channel length and series resistance of the source and drain from the measured drain voltages and currents. The only modification that was made to this method was that many pairs of devices (12 for Device Type 1 and 7 for Device Type 2) were used to permit an estimation of the error in the results. It was determined that the difference ΔL between the drawn gate length L_g and the electrical channel length L was $1.65 \pm 0.04 \mu m$ for Device Type 1 and $0.95 \pm 0.01 \mu m$ for Device Type 2. Since the drawn gate lengths are all known the electrical channel lengths can be easily determined.

The device widths were measured using an optical microscope. They were found to be 49.5 μm for Device Type 1 and 24.8 μm for Device Type 2.

The peak doping in the source and drain regions was determined from four point probe measurements of the sheet resistance in these regions. It was assumed that the doping profile in the source and drain was Gaussian, and then using Irvin's curves [52] N_S was determined for Device Type 1 to be $1.5 \times 10^{20} \text{ cm}^{-3}$ and for Device Type 2 to be $2.3 \times 10^{20} \text{ cm}^{-3}$.

The oxide thickness was found by measuring the capacitance of a large area MOS capacitor in accumulation. The area of the capacitor was measured using an optical microscope and the capacitance of the pad attached to the gate was also determined. The oxide thickness was then found from

$$t_{\text{ox}} = \frac{A \epsilon_{\text{ox}}}{C_{\text{gb}} - C_{\text{pad}}} \quad (3.2.1)$$

to be 854Å for Device Type 1 and 511Å for Device Type 2 where C_{gb} is the capacitance of the MOS capacitor.

The junction depth of the source and drain regions was measured from scanning electron micrographs and found to be 1.03 μm for Device Type 1 [53] and 0.50 μm for Device Type 2 [54].

Determining the flatband voltage posed a problem because the flatband voltage cannot be measured without knowing the substrate doping profile, and the substrate doping profile cannot be

determined without knowing the flatband voltage. To circumvent this conundrum it was decided to calculate the flatband voltage from [72]

$$V_{FB} = - \frac{E_g}{2q} - \phi_{po} - \frac{Q_f}{C_{ox}} \quad (3.2.2)$$

$$\text{where } \phi_{po} = \frac{kT}{q} \ln \left[\frac{N(0)}{n_i} \right] \quad (3.2.3)$$

and where E_g is the bandgap of silicon and $N(0)$ is the doping level at the silicon surface. Note that ϕ_{po} depends upon the doping level at the surface $N(0)$ which is virtually impossible to measure. Fortunately, a reasonable guess at $N(0)$ will suffice because ϕ_{po} has a logarithmic dependence on $N(0)$. Thus for Device Types 1 and 2 $N(0)$ was replaced in (3.2.3) with the doping of the uniform region N_B . For Device Types 1 and 2 this incurs an error in V_{FB} of less than 15 mV.

Equation (3.2.2) assumes that the Fermi level in the polysilicon gate is at the conduction band edge, because the gate is doped heavily n-type. To calculate V_{FB} from (3.2.2) a value for the fixed charge density at the silicon-silicon dioxide interface Q_f must be assumed. A reasonable value to choose for Q_f , since the oxide is well annealed, is $2 \times 10^{10} \text{ q/cm}^2$ based on the results of Akinwande, Ho and Plummer [73]. Using this value for Q_f the flatband voltage is easily calculated from (3.2.2) and is approximately correct. But it was found that for the very exact characterization necessary to verify the subthreshold

current model (3.2.2) was not exact enough. The discrepancies that were encountered were most likely due [74] to the $E_g/(2q)$ term since in the heavily doped polysilicon gate the bandgap is reduced from its usual value of 1.12 eV and the Fermi level is not exactly at the conduction band edge. Despite these drawbacks (3.2.2) was used because the only other alternative of using tabulated values for the work function difference between an n^+ polysilicon gate and a p-type substrate [75], would have produced equally unsatisfactory results. For Device Type 1 the errors inherent in (3.2.2) were compensated for by a slightly perturbed substrate doping profile near the surface. For Device Type 2 the compensation was achieved by reducing the magnitude of V_{FB} predicted using (3.2.2) by 0.17 V, which is approximately the correct magnitude for the bandgap narrowing effect and the difference between the Fermi level and conduction band edge. The issue of dealing with the approximate nature of (3.2.2) is discussed at greater length later in this chapter.

The substrate doping profile $N(x)$ is by far the most difficult device parameter to measure, and yet the subthreshold current is a very sensitive function of it, so it must be measured accurately. The primary method used to measure $N(x)$ was an unpublished technique originally proposed by D. Vincent and then developed by J. Kolk [55] that has been extensively used at Northern Telecom Electronics. The method, which will hereafter be referred to as Kolk's method, uses a long-channel MOSFET and depends upon measuring the AC small signal parameter dV_{SB}/dV_{GS} at a constant drain current. It is capable of measuring the

substrate doping profile from a depth slightly shallower than the zero substrate bias depletion depth $(2\phi_p/M)^{1/2}$, down to the depletion depth at which V_{SB} breaks down the source-substrate junction $[(2\phi_p + V_{SBmax})/M]^{1/2}$ (see (2.1.46) for definition of M). It is extended in this thesis so that it will reveal information about the blind region near the surface. Because a detailed discussion of Kolk's method is unavailable elsewhere it is provided here.

3.2b Measurement of the Substrate Doping Profile

To derive Kolk's method start with the drain current transport equation

$$I_D = - \mu_n W Q_m \frac{d\psi_s}{dy} + \frac{kT}{q} \mu_n W \frac{dQ_m}{dy} \quad (3.2.4)$$

where ψ_s is the surface potential. If the MOSFET is biased so that the gate voltage is just above threshold and the drain voltage V_{DS} is very small, say 0.05 V, then the diffusion term is negligible and the electric field $-d\psi_s/dy$ is nearly constant throughout the channel. If in addition the field dependence of the mobility μ_n is neglected and the device is assumed to be of constant width, then the drain current is equal to a constant K times the mobile charge density Q_m which is nearly constant throughout the channel because of the small drain voltage. Hence

$$I_D = K Q_m \quad (3.2.5)$$

$$\text{where } K = - \mu_n W \frac{d\psi_s}{dy} \quad (3.2.6)$$

Thus under these bias conditions holding the drain current constant ensures that the mobile charge density throughout the channel is held constant.

By charge neutrality at any point in the channel

$$Q_G + Q_D + Q_m = 0 \quad (3.2.7)$$

where Q_G is the charge per unit area on the gate and Q_D is the charge per unit area in the depletion region. If Q_m is held constant by holding I_D constant then

$$dQ_G + dQ_D = 0 \quad (3.2.8)$$

The gate charge is given by

$$Q_G = C_{ox} (V_{GS} - V_{FB} + V_{SB} - \psi_s) \quad (3.2.9)$$

$$\text{thus } dQ_G = C_{ox} (dV_{GS} + dV_{SB} - d\psi_s) \quad (3.2.10)$$

If all potentials are referenced to a zero electrostatic potential in the neutral bulk, then just above threshold the surface potential is almost exactly given by

$$\psi_s = 2\phi_p + V_{SB} \quad (3.2.11)$$

$$\text{where } \phi_p = \frac{kT}{q} \ln \left[\frac{N(x_d)}{n_i} \right] \quad (3.2.12)$$

where x_d is the depletion depth. Since ϕ_p is a very slowly varying function of x_d because of the logarithm, it may be treated as a constant. This allows

$$d\psi_s = dV_{SB} \quad (3.2.13)$$

to be obtained from (3.2.11). Substituting (3.2.13) into (3.2.10), and (3.2.10) into (3.2.8) gives

$$C_{ox} dV_{GS} + dQ_D = 0 \quad (3.2.14)$$

The depletion charge is related to the depletion capacitance C_D by

$$C_D = - \frac{dQ_D}{d\psi_s} \quad (3.2.15)$$

Substituting (3.2.13) into (3.2.15), rearranging and substituting into (3.2.14) gives

$$C_D dV_{SB} = C_{ox} dV_{GS} \quad (3.2.16)$$

The capacitance C_{ox} and C_D are given by

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad (3.2.17)$$

$$C_D = \frac{\epsilon_s}{x_d} \quad (3.2.18)$$

Substituting these two expressions into (3.2.16) gives

$$x_d = R t_{ox} \frac{dV_{SB}}{dV_{GS}} \quad (3.2.19)$$

$$\text{where } R = \frac{\epsilon_s}{\epsilon_{ox}} \quad (3.2.20)$$

The small signal AC parameter dV_{SB}/dV_{GS} at a constant drain current is measured using the circuit shown in Figure 3.2.1. This circuit uses an operational amplifier to adjust V_{SB} so it maintains the constant drain current demanded by the current source as the gate voltage V_{GS} is swept. The AC parameter dV_{SB}/dV_{GS} is then measured by applying a small AC signal to the gate and detecting the AC signal on the substrate at the same frequency using a synchronous detector. Errors due to noise are largely eliminated by adding a low pass filter after the synchronous detector. The AC parameter dV_{SB}/dV_{GS} is then proportional to the RMS value of the AC signal on the output V_x . To set the proportionality the MOSFET and constant current source are replaced by two resistors of known value to produce the circuit shown in Figure 3.2.2. The gain of the op-amp will then be R_2/R_1 which is the same as having an AC parameter dV_{SB}/dV_{GS} equal to R_2/R_1 . The amplitude of the AC input is then set to give a certain proportionality constant between dV_{SB}/dV_{GS} and V_x . For the circuit used in this work the proportionality constant was chosen to be 1/2 so

$$V_x = \frac{1}{2} \frac{dV_{SB}}{dV_{GS}} \quad (3.2.21)$$

Figure 3.2.1. Circuit used by Kolk's method to measure dV_{SB}/dV_{GS} .

Figure 3.2.2. Calibration configuration for the circuit in Figure 3.2.1.

Substituting (3.2.21) into (3.2.19) gives the final expression for the depletion depth x_d as

$$x_d = 2 R t_{ox} V_x \quad (3.2.22)$$

Thus x_d is obtained directly by the measurement of V_x .

The substrate doping at this depletion depth $N(x_d)$ is determined by noting that

$$dQ_D = -q N(x_d) dx_d \quad (3.2.23)$$

Taking differentials of (3.2.22) to eliminate dx_d from (3.2.23), and eliminating dQ_D using (3.2.14), yields the following expression for the substrate doping as a function of depletion depth

$$N(x_d) = \frac{C_{ox}^2}{2 q \epsilon_s} \frac{dV_{GS}}{dV_x} \quad (3.2.24)$$

Note that the oxide thickness must be supplied by an extra measurement to use equations (3.2.22) and (3.2.24). Measuring the oxide capacitance and area of a large MOS capacitor, and then using equation (3.2.1) is a good technique to do this.

A convenient way to implement Kolk's method is to use the circuit in Figure 3.2.1 in conjunction with a HP4145 Semiconductor Parameter Analyzer. The 4145 is used to set the gate voltage and provide the constant current source, as well as monitor V_{SB} and V_x . The derivative dV_{GS}/dV_x can be taken numerically by the 4145 thereby permitting the calculation of

$N(x_d)$ from (3.2.24). The substrate doping profile $N(x_d)$ versus x_d can then be plotted on the screen of the 4145 or on an x-y recorder.

To use Kolk's method to profile at depths shallower than the zero substrate bias depletion depth $(2\phi_p/M)^{1/2}$, it is necessary to forward bias the source-substrate junction. But it is found in practice that it is possible to go only slightly shallower than $(2\phi_p/M)^{1/2}$. This is because when more than 0.3-0.4 V of forward bias is placed on the source-substrate junction the current that is conducted shunts a significant amount of current away from the channel, thereby making the drain current no longer constant and Kolk's method impractical.

The traditional MOS capacitor profiling method originated by Van Gelder and Nicollian [56] can measure the substrate doping profile from three Debye lengths below the surface (i.e. $3L_D$) to $(2\phi_p/M)^{1/2}$. The substrate doping profile is calculated using this method from

$$N(x_d) = -2 \left[q \epsilon_s \frac{d}{dv_{GB}} \left(\frac{1}{C_{gb}^2} \right) \right]^{-1} \quad (3.2.25)$$

whose derivation is given in [56]. Thus if a large area MOS capacitor is available near the MOSFET used for Kolk's method, then the method of Van Gelder and Nicollian can be used to extend the profiling to within $3L_D$ of the surface. Kolk's method is thus an alternative to the pulsed capacitance method [56] which

is normally used to extend the profiling depth of the conventional MOS capacitor technique.

The $3L_D$ minimum depth limitation is caused by the fact that for depletion depths less than $3L_D$ the majority carrier density is no longer equal to the substrate doping density because it has an exponential dependence on the surface potential. To partially overcome this limitation Kolk's method can be used to obtain the dose D_I and centroid x_C of the nonuniformity in the profile from the surface down to a region which is fairly uniform. As explained below, it is possible to use this dose and centroid to deduce the doping profile in the blind region which extends from the surface to $3L_D$, or, if the MOS capacitor technique is not used, then it extends from the surface to approximately $(2\phi_p/M)^{1/2}$.

The simplest way to obtain D_I and x_C using Kolk's method is to measure $N(x_d)$ using two different constant drain currents I_{D1} and I_{D2} where $I_{D2} = 2 I_{D1}$. Since Q_m is proportional to I_D by (3.2.5) the V_{GS} versus x_d curves for I_{D1} and I_{D2} can be extrapolated to an $I_D = 0$ curve for which $Q_m = 0$ (see Figure 3.2.3). In practice all that really needs to be extrapolated is the V_{GS} axis intercept of the linear portion of the V_{GS} versus x_d curves that corresponds to the uniformly doped region. The fact that the V_{GS} versus x_d curve is linear when the depletion edge is in a uniformly doped region is obvious from equation (3.2.31) which is derived below. By virtue of the fact

	GRAD	1/GRAD	Xintercept	Yintercept
LINE1	4.36E+00	229E-03	167E-03	-727E-03
LINE2	4.45E+00	225E-03	158E-03	-702E-03

N () = 4.8882E15 • $\Delta V_G / \Delta V_W$
W (um) = .5187 • V_W

Figure 3.2.3. V_{GS} versus x_d curves for $I_D = I_{D1}$, $I_D = I_{D2}$ and $I_D = 0$.

that $I_{D2} = 2 I_{D1}$ the V_{GS} intercept for the $I_D = 0$ curves is given by

$$b_0 = 2 b_1 - b_2 \quad (3.2.26)$$

where b_1 and b_2 are respectively the intercepts of the linear portion of the I_{D1} and I_{D2} curves.

To derive an equation for D_I (3.2.9) is substituted into (3.2.7) and Q_m is set to zero. Then since

$$Q_{D1} = -q \int_0^{x_d} N(x) dx \quad (3.2.27)$$

it follows that

$$C_{ox} (V_{GS} - V_{FB} + V_{SB} - \psi_s) + q \int_0^{x_d} N(x) dx = 0 \quad (3.2.28)$$

Now if x_d is in the uniformly doped region then

$$\int_0^{x_d} N(x) dx = N_B x_d + D_I \quad (3.2.29)$$

$$\text{since } D_I \text{ is defined by } D_I = \int_0^{x_d} [N(x) - N_B] dx \quad (3.2.30)$$

where N_B is the doping in the uniform region and D_I is the dose of the nonuniformity from the surface down to the uniform region. Substituting (3.2.29) and (3.2.11) into (3.2.28) and rearranging yields

$$V_{GS} = \frac{q N_B}{C_{ox}} x_d + V_{FB} + \frac{q D_I}{C_{ox}} + 2\phi_p \quad (3.2.31)$$

From (3.2.31) the intercept of the linear portion of the V_{GS} versus x_d for $Q_m = 0$ curve is

$$b_0 = V_{FB} + \frac{q D_I}{C_{ox}} + 2\phi_p \quad (3.2.32)$$

Upon rearranging, the dose of the nonuniformity D_I is given by

$$D_I = \frac{C_{ox}}{q} (b_0 - V_{FB} - 2\phi_p) \quad (3.2.33)$$

$$\text{where here } \phi_p = \frac{kT}{q} \ln \left(\frac{N_B}{n_i} \right) \quad (3.2.34)$$

The flatband voltage can be calculated from (3.2.2) and then D_I is found from (3.2.33).

The centroid x_C is related to the doping profile by

$$x_C = \frac{1}{D_I} \int_0^{x_d} x [N(x) - N_B] dx \quad (3.2.35)$$

According to [57] the depletion depth is related to the surface potential for a nonuniform substrate whose nonuniformity is entirely depleted by

$$x_d = \sqrt{2} L_D \left(B\psi_s - \frac{D_I x_C}{N_B L_D^2} - 1 \right)^{1/2} \quad (3.2.36)$$

$$\text{where } L_D = \sqrt{\frac{kT}{q} \frac{\epsilon_s}{q N_B}} \quad (3.2.37)$$

Rearranging (3.2.36) to

$$D_I x_C = \frac{N_B}{2} [2 L_D^2 (\beta \psi_s - 1) - x_d^2] \quad (3.2.38)$$

allows $D_I x_C$ to be calculated from ψ_s . Previously, the approximate value of $2\phi_p + V_{SB}$ has been used for the surface potential ψ_s (see (3.2.11)), and it is possible to calculate $D_I x_C$ from (3.2.38) using this value. But a more accurate value of $D_I x_C$ can be obtained by calculating ψ_s from (3.2.39) which is derived by substituting (3.2.29) into (3.2.27), and (3.2.27) and (3.2.9) into (3.2.7). This yields

$$C_{ox} (V_{GS} - V_{FB} + V_{SB} - \psi_s) - q N_B x_d - q D_I + Q_m = 0 \quad (3.2.39)$$

Q_m is found by neglecting the diffusion current in (3.2.4) and approximating $d\psi_s/dy$ as V_{DS}/L . Substituting the resulting expression into (3.2.39) gives

$$C_{ox} (V_{GS} - V_{FB} + V_{SB} - \psi_s) - q N_B x_d - q D_I - \frac{I_D L}{\mu_n W V_{DS}} = 0 \quad (3.2.40)$$

Solving (3.2.40) for ψ_s produces

$$\psi_s = V_{GS} - V_{FB} + V_{SB} - \frac{1}{C_{ox}} (q N_B x_d + q D_I + \frac{I_D L}{\mu_n W V_{DS}}) \quad (3.2.41)$$

where the mobility is calculated from the gate field dependent part of Coen and Muller's [58] mobility model

$$\mu_n = \frac{\mu_{no}}{1 + \frac{V_{GS} - V_{FB} + V_{SB} - \psi_s}{R t_{ox} E_{Cx}}} \quad (3.2.42)$$

where μ_{no} is $850 \text{ cm}^2/\text{Vs}$ and E_{Cx} is $7 \times 10^5 \text{ V/cm}$ for (100) p-type silicon. Equations (3.2.41) and (3.2.42) are weakly transcendental in ψ_s and can be used to calculate ψ_s in an iterative fashion for each set of values of V_{GS} , V_{SB} , x_d and I_D . A good initial guess for ψ_s is $2\phi_p + V_{SB}$. Once ψ_s is determined it is substituted into (3.2.38) to calculate $D_I x_C$. Since a value of $D_I x_C$ can be determined for every value of V_{GS} used to measure x_d and $N(x_d)$, the most accurate determination of $D_I x_C$ is obtained by taking a mean and standard deviation from all the calculated results. Dividing the mean value of $D_I x_C$ by D_I obtained from (3.2.33) yields x_C . The way in which D_I and x_C are used to determine the substrate doping profile is discussed subsequently.

Kolk's method was used to measure the doping profile for Device Types 1 and 2, and for Device Type 2 the MOS capacitor profiling technique was also used. These results are presented in Figures 3.2.4 and 3.2.5. Note that the data from Kolk's method and the MOS capacitor method agree in the region in which they overlap. This confirms the accuracy of Kolk's method. After averaging the doping profile in the uniform region to

Figure 3.2.4. Measured doping profile for Device Type 1.

Figure 3.2.5. Measured doping profile for Device Type 2.

obtain N_B , the technique for calculating D_I and x_C was used. The results are given in Table 3.2.1.

Table 3.2.1. Measured values of N_B , $D_I x_C$ and D_I and x_C for Device Types 1 and 2.

	N_B (cm^{-3})	$D_I x_C$ (cm^{-1})	D_I (cm^{-2})	x_C (μm)
Device Type 1	1.106×10^{16}	$-3.80 \pm 0.3 \times 10^6$	-1.3944×10^{11}	0.27252
Device Type 2	2.094×10^{16}	$-3.87 \pm 0.3 \times 10^6$	-1.5594×10^{11}	0.24817

Device Types 1 and 2 are both fabricated in a deep p-well that is approximately uniformly doped in the vicinity of $1 \mu\text{m}$ below the surface, but near the surface the doping is depleted due to boron segregation into the oxide. In addition, Device Type 1 has a shallow threshold adjust implant of boron. If the constant doping level is taken to be N_B , and the boron segregation and shallow implant are both modelled as Gaussians centered on the surface, then the substrate doping profiles can be expressed as

$$N(x) = N_B + N_1 e^{-(x/\sigma_1)^2} + N_2 e^{-(x/\sigma_2)^2} \quad (3.2.43)$$

for both devices.

For Device Type 1 N_1 and σ_1 were determined by fitting to the measured profile in a region deep enough to be unaffected by the surface implant. N_2 and σ_2 were then determined from

$$\sigma_2 = \frac{N_1 \sigma_1^2 + 2 D_I x_C}{\sqrt{\pi} + N_1 \sigma_1} \quad (3.2.44)$$

$$N_2 = \frac{2 D_I}{\sigma_2} + N_1 \sigma_1 \quad (3.2.45)$$

which are derived by substituting (3.2.43) into the definitions of D_I and x_C (3.2.30) and (3.2.35), performing the integrals, and solving the two resulting equations for N_2 and σ_2 . It was found that to obtain reasonable agreement between measured and simulated subthreshold currents $D_I x_C$ had to be adjusted to $-4.45 \times 10^6 \text{ cm}^{-1}$. This is 2.2 standard deviations away from the mean listed in Table 3.2.1. With this value of $D_I x_C$ the following substrate doping profile was obtained for Device Type 1.

$$N(x) = 1.106 \times 10^{16} - 1.042 \times 10^{16} e^{-(x/0.33)^2} + 1.4214 \times 10^{16} e^{-(x/0.13122)^2} \quad (3.2.46)$$

where x is microns and $N(x)$ is in cm^{-3} . This profile is shown in Figure 3.2.4 and is in excellent agreement with the measured data where it is available. Notice that by using the dose D_I and centroid x_C an approximate shape for the surface implant has been determined where otherwise it would have remained entirely unmeasured.

Since there is no threshold implant for Device Type 2 the second Gaussian is neglected. The values of N_1 and σ_1 were determined by fitting to the measured data giving the following profile.

$$N(x) = 2.094 \times 10^{16} - 9.066 \times 10^{15} e^{-(x/0.27606)^2} \quad (3.2.47)$$

The values of $D_I x_C$, D_I and x_C are calculated from this profile using

$$D_I x_C = \frac{N_1 \sigma_1^2}{2} \quad (3.2.48)$$

$$D_I = \frac{\sqrt{\pi}}{2} N_1 \sigma_1 \quad (3.2.49)$$

$$x_C = \frac{\sigma_1}{\sqrt{\pi}} \quad (3.2.50)$$

which are calculated by substituting (3.2.43) into (3.2.30) and (3.2.35), setting N_2 to zero, and performing the integrals. Using the values of N_1 and σ_1 specified in (3.2.47) these three equations give $D_I x_C = -3.45 \times 10^6 \text{ cm}^{-1}$, $D_I = -2.2180 \times 10^{11} \text{ cm}^{-2}$ and $x_C = 0.15575 \text{ } \mu\text{m}$. This value of $D_I x_C$ agrees reasonably well

with the measured value listed in Table 3.2.1 as it is 1.4 standard deviations from the mean, but the values of D_I and x_C do not agree with measured values. The errors appear to be related to the fact that unless the magnitude of the flatband voltage given by (3.2.2) is reduced by 0.17 V, good agreement between measured and simulated subthreshold currents cannot be obtained. In other words the flatband voltage is now given by

$$V_{FB} = -\frac{E_g}{2q} - \phi_{po} - \frac{Q_f}{C_{ox}} + \Delta V_{FB} \quad (3.2.51)$$

where $\Delta V_{FB} = 0.17$ V. Using this value of V_{FB} changes the measured values of D_I and x_C because of the presence of V_{FB} in (3.2.33). The measured values are now $D_I = -2.2769 \times 10^{11} \text{ cm}^{-2}$ and $x_C = 0.16997 \text{ } \mu\text{m}$ that agree well with the values obtained from the measured profile given by (3.2.47) which are $D_I = -2.2180 \times 10^{11} \text{ cm}^{-2}$ and $x_C = 0.15575 \text{ } \mu\text{m}$. Thus with the use of V_{FB} as specified by (3.2.51) the measurements obtained by Kolk's method for Device Type 2 are now all consistent.

The flatband voltage had to be modified for Device Type 2 because of the approximate nature of equation (3.2.2). As mentioned previously, what makes (3.2.2) approximate is the term $-E_g/(2q)$ because it assumes that the Fermi level in the n^+ polysilicon gate is at the conduction band edge. It is almost certainly lower than this and there is almost certainly a reduced value of the bandgap E_g in the heavily doped polysilicon gate, both of which are consistent with the fact that a positive value of ΔV_{FB} had to be used in (3.2.51). Clearly the value of V_{FB}

used for Device Type 1 is also in error but this has been compensated for in the dose of the threshold implant, which, because it is close to the surface, has almost the same effect. What is needed to avoid these problems and have unambiguous characterization of the devices is a method to independently measure V_{FB} . The subthreshold current model could be used for this purpose but because this is the very thing that is being tested its use for this purpose is ruled out.

When the subthreshold current model described in Section 2.4 was used with the substrate doping profiles given by (3.2.46) and (3.2.47) for Device Types 1 and 2 respectively, good agreement was obtained between measured and simulated subthreshold currents for the short-channel devices. These subthreshold characteristics are displayed in Figures 3.2.6 and 3.2.7. Note in particular the agreement as V_{DS} is increased from 0.1 V to 5 V for the shortest channel length devices because this is the most sensitive test of the ability of the subthreshold model to simulate short-channel effects. Surprisingly, the most difficulty in obtaining agreement between measured and simulated results was experienced for the long-channel devices. The reason for this is explained by observing the ratio of measured subthreshold currents and measured channel lengths displayed in Tables 3.2.2 and 3.2.3.

Figure 3.2.6. Measured and simulated subthreshold I_D versus V_{GS} characteristics for Device Type 1 where $V_{SB} = 0$ and $V_{DS} = 0.1$ V unless otherwise indicated.

Figure 3.2.7. Measured and simulated subthreshold I_D versus V_{GS} characteristics for Device Type 2 where $V_{SB} = 0$ and $V_{DS} = 0.1$ V unless otherwise indicated.

Table 3.2.2. Ratios of drain current and channel length versus channel length with fixed charge values obtained from fitting measured and simulated subthreshold currents for Device Type 1. The bias conditions used for the first two columns were $V_{GS} = 0.5$ V, $V_{DS} = 0.1$ V and $V_{SB} = 0$.

L (μm)	$I_D(L)$ $I_D(23.35 \mu\text{m})$	23.35 μm L	Q_f (q/cm^2)
23.35	1	1	3.4×10^{10}
8.35	2.222	2.796	2.8×10^{10}
6.35	2.826	3.677	2.6×10^{10}
3.35	4.775	6.970	2.0×10^{10}
2.35	9.149	9.936	2.0×10^{10}
1.35	76.797	17.296	2.0×10^{10}

Table 3.2.3. Ratios of drain current and channel length versus channel length with fixed charge values obtained from fitting measured and simulated subthreshold currents for Device Type 2. The bias conditions used for the first two columns were $V_{GS} = 0.5$ V, $V_{DS} = 0.1$ V and $V_{SB} = 0$.

L (μm)	$I_D(L)$ $I_D(17.05 \mu\text{m})$	17.05 μm L	Q_f (q/cm^2)
17.05	1	1	3.0×10^{10}
3.25	4.348	5.246	2.0×10^{10}
2.65	5.676	6.434	2.0×10^{10}
2.05	8.914	8.317	2.0×10^{10}
1.45	14.131	11.759	2.0×10^{10}
1.15	28.049	14.826	2.0×10^{10}
0.85	117.500	20.059	2.0×10^{10}

Since the drain current I_D varies as $1/L$ the ratio of $I_D(L_{\text{short}})/I_D(L_{\text{long}})$ should be equal to $L_{\text{long}}/L_{\text{short}}$ if short-channel effects are negligible, or it should be greater than $L_{\text{long}}/L_{\text{short}}$ if short-channel effects are present. But this is not what is observed in Tables 3.2.2 and 3.2.3. For Device Type 1 it is found that $I_D(L)/I_D(23.25 \text{ } \mu\text{m})$ is less than $23.35 \text{ } \mu\text{m}/L$ for every device except for the $1.35 \text{ } \mu\text{m}$ channel length device where short-channel effects finally swamp the anomalous effect. Similarly for Device Type 2 $I_D(L)/I_D(17.05 \text{ } \mu\text{m})$ is less than $17.05 \text{ } \mu\text{m}/L$ for the $3.25 \text{ } \mu\text{m}$ and $2.65 \text{ } \mu\text{m}$ devices, after which it is greater than $17.05 \text{ } \mu\text{m}/L$ because short-channel effects overwhelm the anomalous effect. This anomalous effect cannot be explained by errors in the channel length measurements because it is too large. It seems unlikely that it is due to variations in the substrate doping profile, though this is conceivable. The most likely cause is that the flatband voltage increases in magnitude for longer channel lengths although the physical reason for this is unknown. If the increase in the flatband voltage is interpreted as an increase in the fixed charge at the silicon surface Q_f (as a possible, but not proven, physical explanation) then the Q_f values listed in Tables 3.2.2 and 3.2.3 produce agreement between measured and simulated subthreshold currents. These Q_f values have a definite pattern and are by no means random. For all channel lengths less than or equal to $3.35 \text{ } \mu\text{m}$ for both Device Types Q_f is $2.0 \times 10^{10} \text{ q/cm}^2$, while for larger channel lengths Q_f increases monotonically with channel length.

Thus it appears that using these Q_f values to modify the flatband voltage is ~~reasonable~~ although the physical explanation for this phenomenon is unknown. When this is done, good agreement between measured and simulated subthreshold currents is obtained for all channel lengths for Device Types 1 and 2 (see Figures 3.2.6 and 3.2.7).

3.2c Triple Agreement

Since it was so difficult to achieve a sufficiently accurate characterization of the devices to permit a reasonable comparison of measured and simulated results, it was decided to try to obtain triple agreement between measurements, analytical model simulations, and MINIMOS simulations to dispel any doubts about the authenticity of the results. This was done for Device Type 1 by plotting I_D versus L for $V_{GS} = 0.5$ V, $V_{DS} = 0.1$ V and $V_{SB} = 0$ for the three sets of data. The only adjustment that was done here was to use a metallurgical junction spacing on the silicon ~~surface~~ that was $0.2\text{ }\mu\text{m}$ less than the measured electrical channel length for the MINIMOS simulations. This was necessary because the measured electrical channel length extends slightly into the source and drain regions so the metallurgical junction spacing at the surface is always less than the channel length. The distance $0.2\text{ }\mu\text{m}$ is comparable to the distance $0.35\text{ }\mu\text{m}$ found by Viswanathan et al. [28] for the same effect but with different devices. The agreement between measured, analytically simulated, and numerically simulated results is shown in Figure 3.2.8. This triple agreement, taken along with the agreement between measured

Figure 3.2.8. Drain current versus channel length for measured, analytically simulated and MINIMOS simulated results using Device Type 1 where $V_{GS} = 0.5 V$, $V_{DS} = 0.1 V$ and $V_{SB} = 0$ unless otherwise indicated.

and simulated results displayed in Figures 3.2.6 and 3.2.7, confirms that the short-channel subthreshold current model can accurately predict the subthreshold behavior of a short-channel MOSFET solely from its device structure and applied biases, without the use of fitting parameters.

CHAPTER 4

SOLUTION OF THE POISSON PROBLEM FOR THE GENERAL MODEL

4.1 The Poisson Problem and its Boundary Conditions

The subthreshold model is useful and demonstrates the accuracy and potential of the two-dimensional analytical approach; but it is not a complete description of the behavior of the short-channel MOSFET because it is not valid in the triode and saturation regions. As discussed in Chapter 1, to derive a general model, valid in all these regimes of operation, it is necessary to simultaneously solve a two-dimensional Poisson's equation and a one-dimensional current continuity equation for the geometry and boundary conditions of the MOSFET. The use of a one-, as opposed to a two-dimensional current continuity equation, will give accurate results provided that there is surface conduction along at least part of the channel, and where there is bulk conduction the mobile charge concentration is less than or approximately equal to the bulk doping. This simplification is necessary to make the problem analytically tractable.

Introducing even a one-dimensional current continuity equation into the problem greatly increases its analytical complexity. Because of this it is necessary to assume that the substrate is uniformly doped, thereby sufficiently reducing the complexity of the problem so that there is a reasonable hope of finding an analytical solution. Once the solution for a uniform

substrate device is found it can be used as a basis to determine the solution for a device with an arbitrary substrate doping profile. An extension of this type is discussed in Chapter 8. For a device with a uniformly doped substrate the two-dimensional Poisson's equation, its associated boundary conditions, and the one-dimensional current continuity equation, are formulated as a Poisson problem on the cross-hatched rectangular box of depth d shown in Figure 4.1.1 as follows:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{qN_B}{\epsilon_s} = K \quad (4.1.1)$$

$$\psi(x, 0) = f_S(x) \quad (4.1.2)$$

$$\psi(x, L) = f_D(x) \quad (4.1.3)$$

$$C_{ox} [V_{GB} - \psi(0, y)] + Q_m(y) + \epsilon_s \frac{\partial \psi}{\partial x}(0, y) = 0 \quad (4.1.4)$$

$$\frac{\partial \psi}{\partial x}(d, y) = \frac{\partial U}{\partial x}(d, y) \quad (4.1.5)$$

$$I_D = \mu_n W \left[-Q_m \frac{\partial \psi}{\partial y}(0, y) + \frac{1}{B} \frac{dQ_m}{dy} \right] \quad (4.1.6)$$

$$\frac{dI_D}{dy} = 0 \quad (4.1.7)$$

where $\psi(x, y)$ is the two-dimensional electrostatic potential and $U(x, y)$ is the long-channel trial function, and both are referenced to a zero potential in the neutral substrate. This is the Poisson problem that must be solved to derive the general

Figure 4.1.1. The MOSFET geometry with the rectangular domain for the general solution indicated by cross-hatching.

model of a short-channel MOSFET with a uniformly doped substrate. Equation (4.1.1) is Poisson's equation in the depletion region of a uniformly doped substrate. The next two equations, (4.1.2) and (4.1.3), are the boundary conditions on the sides of the rectangular region under consideration (see Figure 4.1.1). They use the as yet unspecified functions $f_S(x)$ and $f_D(x)$. These functions are derived later in this chapter and take the form of either one or two parabolas depending on the bias conditions. The functions $f_S(x)$ and $f_D(x)$ make the side boundary conditions more accurate than the cubic functions derived in Chapter 2. Equation (4.1.4) is the surface boundary condition. It is a statement of the continuity of normal displacement across the silicon-silicon dioxide interface. It assumes that all the mobile charge is located in an infinitely thin sheet at the interface of charge density Q_m .

The boundary condition on the bottom of the rectangular box is (4.1.5). It neglects the short-channel component of electric displacement crossing the line $x = d$ since $U(x,y)$ is the long-channel trial function. This is the same approximation made by (2.1.7) in the subthreshold model. The difference between (2.1.7) and (4.1.5) is that in the subthreshold model d is the true minimum depletion depth while in the general model d is made equal to the minimum depletion depth of the ideal long-channel solution. Hence:

$$d = \sqrt{\frac{2}{K} U(0,0)} \quad (4.1.8)$$

It is necessary to use this value for d because for greater depths the trial function $U(x,y)$ used in the general model is not valid. A consequence of employing the value specified by (4.1.8) for d is that the bottom of the rectangular region under consideration does not touch the depletion region edge, as is shown in Figure 4.1.1.

The last two equations, (4.1.6) and (4.1.7), are a statement of one-dimensional current continuity. Equation (4.1.6) is the drain current transport equation and (4.1.7) is the actual statement of current continuity. The role of these two equations in the Poisson problem is to complete the surface boundary condition (4.1.4). They do this by specifying how the mobile charge should arrange itself on the silicon surface to produce current continuity. This determines the function $Q_m(y)$, and completes boundary condition (4.1.4).

4.2 The Side Boundary Conditions

To finish the description of the Poisson problem the side boundary condition functions $f_S(x)$ and $f_D(x)$ must be derived. They are made more accurate than the cubic boundary conditions given by (2.1.39) and (2.1.40), by using the fact that the normal electric displacement is continuous at points $(0,0)$ and $(0,L)$. Taking this fact into account in the derivation of $f_S(x)$ and $f_D(x)$ automatically satisfies (4.1.4) at $(0,0)$ and $(0,L)$, which is necessary for the physical accuracy and mathematical stability of the general Poisson solution.

It has been found from MINIMOS simulations that if there is a significant amount of mobile charge at points $(0,0)$ and $(0,L)$, then the potentials $\psi(x,0)$ and $\psi(x,L)$ resemble parabolas, as shown in Figure 4.2.1a. But if the mobile charge is negligible, say at the point $(0,L)$, then the gate pulls down the potential in a surface region that extends from the surface to a depth x_m , while the drain pulls up the potential below this depth, as shown in Figure 4.2.1b. The resulting potential function $\psi(x,L)$ can be modelled as two parabolas, one in the surface region and one in the bulk region, that are matched at depth x_m , by demanding that the potential and electric field be continuous.

The single parabola boundary condition takes the form

$$\psi(x) = \psi_s (1 - x/x_0)^2 \quad (4.2.1)$$

where $\psi(x)$, ψ_s and x_0 can be respectively either $\psi(x,0)$, $\psi(0,0)$ and x_s at the source, or $\psi(x,L)$, $\psi(0,L)$ and x_D at the drain. This form of (4.2.1) has been chosen so that it satisfies the zero field and zero potential conditions at the depletion edge where $x = x_0$. To complete the boundary condition the surface potential ψ_s is calculated so that normal displacement is continuous at the surface. Hence it must satisfy

$$C_{ox} (V_{GB} - \psi_s) + Q_m(\psi_s, \phi_n) + \epsilon_s \left. \frac{\partial \psi}{\partial x} \right|_{x=0} = 0 \quad (4.2.2)$$

where $Q_m(\psi_s, \phi_n)$ is given by the sheet charge expression for the mobile charge density

(a)

(b)

Figure 4.2.1. Potential profile $\psi(x,L)$ extracted from the results of a MINIMOS simulation for a device whose parameters are: $L = 1 \text{ } \mu\text{m}$, $W = 24 \text{ } \mu\text{m}$, $N_B = 2 \times 10^{16} \text{ cm}^{-3}$, $N_S = 2.3 \times 10^{20} \text{ cm}^{-3}$, $t_{ox} = 500 \text{ \AA}$, $r_j = 0.532 \text{ } \mu\text{m}$, $V_{FB} = -0.97152 \text{ V}$ where for (a) $V_{GS} = 5 \text{ V}$, $V_{DS} = 0.1 \text{ V}$, $V_{SB} = 0$ and for (b) $V_{GS} = 1 \text{ V}$, $V_{DS} = 5 \text{ V}$, $V_{SB} = 0$.

$$Q_m(\psi_s, \phi_n) = -C_K \{ [e^{\beta(\psi_s - \phi)} + \beta\psi_s - 1]^{1/2} - (\beta\psi_s - 1)^{1/2} \} \quad (4.2.3)$$

$$\text{where } C_K = \sqrt{\frac{2 q \epsilon_s N_B}{\beta}} \quad (4.2.4)$$

$$\phi = \phi_n + \phi_p \quad (4.2.5)$$

$$\text{and where } \phi_n(0) = \phi_p + V_{SB} \quad (4.2.6)$$

$$\phi_n(L) = \phi_p + V_{SB} + V_{DS} \quad (4.2.7)$$

are the values of the quasi-Fermi potential ϕ_n at the source and drain respectively.

$$\text{From (4.2.1) } \left. \frac{\partial \psi}{\partial x} \right|_{x=0} = - \frac{2 \psi_s}{x_0} \quad (4.2.8)$$

Substituting (4.2.3) and (4.2.8) into (4.2.2) yields

$$C_{ox} (V_{GB} - \psi_s) - \frac{2 \epsilon_s}{x_0} \psi_s - C_K \{ [e^{\beta(\psi_s - \phi)} + \beta\psi_s - 1]^{1/2} - (\beta\psi_s - 1)^{1/2} \} = 0 \quad (4.2.9)$$

Upon differentiating (4.2.9) the following Newton-Raphson equation in ψ_s is obtained

$$\begin{aligned}
\psi'_S = \psi_S + [C_{ox} (V_{GB}' - \psi_S) - \frac{2 \epsilon_S}{x_0} \psi_S \\
- C_K ([e^{\beta(\psi_S - \phi)} + \beta\psi_S - 1]^{1/2} - (\beta\psi_S - 1)^{1/2})] / [C_{ox} + \frac{2 \epsilon_S}{x_0} \\
+ \frac{\beta}{2} C_K \{ \frac{e^{\beta(\psi_S - \phi)}}{[e^{\beta(\psi_S - \phi)} + \beta\psi_S - 1]} + \frac{1}{(\beta\psi_S - 1)^{1/2}} \}] \quad (4.2.10)
\end{aligned}$$

where ψ'_S is the next value in the iterative sequence.

To obtain good initial guesses for use with (4.2.10) a threshold voltage V_{GBT} and a critical gate voltage V_{GBCr} must be defined. Threshold occurs when $\psi_S = \phi$. Substituting this condition into (4.2.9), making the approximation that $(\beta\psi_S - 1)^{1/2} \approx (\beta\psi_S)^{1/2}$, and rearranging gives

$$V_{GBT}' = \phi (1 + \frac{2 R t_{ox}}{x_0}) \quad (4.2.11)$$

The critical gate voltage V_{GBCr}' delineates the regime in which the mobile charge is negligible from the regime in which it is not. The boundary between the two regimes is here defined to be when the surface potential is $2 kT/q$ below the value at threshold. Hence the critical surface potential is defined to be

$$\psi_{Cr} = \phi - 2 \frac{kT}{q} \quad (4.2.12)$$

To obtain the expression for V'_{GBCr} substitute (4.2.12) into (4.2.9), neglect the electron charge term, and rearrange to give

$$V'_{GBCr} = \psi_{Cr} \left(1 + \frac{2 R t_{ox}}{x_0} \right) \quad (4.2.13)$$

The importance of V'_{GBCr} is that for gate voltages larger than it the mobile charge is nonnegligible and the single parabola boundary condition is used, while for gate voltages less than V'_{GBCr} the mobile charge is negligible and it is necessary to use the double parabola boundary condition.

Above threshold (i.e. for $V'_{GB} > V'_{GBT}$), the initial guess for ψ_s is obtained by first using the crude guess $\psi_s = \phi$ and then improving it by substituting it into (4.2.9) everywhere except the exponential. Upon isolating the ψ_s in the exponential the above threshold initial guess is found to be

$$\psi_s = \phi + \frac{1}{\beta} \ln \left[\left(\frac{1}{C_K} [C_{ox}(V'_{GB} - \phi) - \frac{2 \epsilon_s}{x_0} \phi] + (\beta\phi - 1)^{1/2} \right)^2 - \beta\phi + 1 \right] \quad (4.2.14)$$

For gate voltages below V'_{GBT} but above V'_{GBCr} the initial guess is obtained by neglecting the electron charge term in (4.2.9) and solving for ψ_s . This gives

$$\psi_s = \frac{V'_{GB}}{1 + \frac{2 R t_{ox}}{x_0}} \quad (4.2.15)$$

For V_{GB}' less than V_{GBCr}' no initial guess is required and the double parabola boundary condition developed below is used.

The procedure used to calculate ψ_s is to start with (4.2.14) as an initial guess if $V_{GB}' > V_{GBT}'$, or to use (4.2.15) as an initial guess if $V_{GBCr}' < V_{GB}' < V_{GBT}'$. A maximum of two iterations using (4.2.10) is then needed to calculate ψ_s to seven digit accuracy. Once ψ_s is found the single parabola boundary condition (4.2.1) is completely determined.

Calculating the surface potential for the single parabola boundary condition is just one example of the general problem that involves using the continuity of normal electric displacement at the silicon surface to define the surface potential. This problem is encountered three times in the derivation of the general model: in the calculation of the surface potential for the single parabola boundary condition; in the calculation of potentials $U(0,0)$ and $U(0,L)$ that are needed to define the long-channel trial function $U(x,y)$; and in the enforcement of the continuity of normal electric displacement in the iterative function technique discussed in Chapter 5. The method used to calculate the surface potential is essentially the same for all three applications. First, an equation that defines the surface potential and is based on the continuity of normal electric displacement at the silicon surface is derived (e.g. (4.2.9)). Upon taking the derivative of the defining

equation a Newton-Raphson equation in terms of the surface potential is obtained (e.g. (4.2.10)). To facilitate rapid calculation of the surface potential using the Newton-Raphson expression two initial guesses are required: one above threshold and one below threshold (e.g. (4.2.14) and (4.2.15)). The reason for this is that the defining equation has a radically different behavior above threshold where the electron charge term is dominant, and below threshold where the electron charge term is negligible. To choose between the two initial guesses a threshold voltage is used to decide whether the point in question is above or below threshold. The need to have a critical gate voltage V_{GBCr} is peculiar to the calculation of the surface potential for the single parabola boundary condition, because if V_{GB} is less than V_{GBCr} then the calculation of the surface potential is unnecessary as the double parabola boundary condition is used. Since the derivations needed to calculate the surface potentials for the trial function $U(x,y)$ and the iterative function technique follow the above pattern, their descriptions below will be kept to a minimum in order to avoid repetition.

It has been found using MINIMOS simulations that on the interval $[x_m, x_0]$ the side boundary conditions are well described by

$$\psi(x) = \psi_{cr} (1 - x/x_0)^2 \quad (4.2.16)$$

when $V_{GB}' \leq V_{GBCr}'$. In other words once V_{GB}' falls below V_{GBCr}' the potential below x_m does not change significantly. This then defines one of the two parabolas. The second parabola is written as

$$\psi(x) = a_0 + a_1 x + a_2 x^2 \quad (4.2.17)$$

The three coefficients a_0 , a_1 and a_2 are calculated from the three conditions that potential and field must be continuous at depth x_m , and that the normal displacement must be continuous at the surface. Solving the three equations in three unknowns generated by these conditions yields

$$a_0 = \frac{V_{GB}' x_m / (R t_{ox}) + 2 \psi_m - \psi_m' x_m}{2 + x_m / (R t_{ox})} \quad (4.2.18)$$

$$a_1 = \frac{a_0 - V_{GB}'}{R t_{ox}} \quad (4.2.19)$$

$$a_2 = \frac{\psi_m' - a_1}{2 x_m} \quad (4.2.20)$$

$$\text{where } \psi_m = \psi_{Cr} (1 - x_m/x_0)^2 \quad (4.2.21)$$

$$\psi_m' = - \frac{2 \psi_{Cr}}{x_0} (1 - x_m/x_0) \quad (4.2.22)$$

are the potential and its first derivative at x_m .

Equations (4.2.1) to (4.2.15) define the single parabola boundary condition and equations (4.2.16) to (4.2.22) define the double parabola boundary condition, so the functions $f_S(x)$ and $f_D(x)$ can now be written as

$$\text{If } V_{GB} > V_{GBCrS} \quad f_S(x) = \psi(0,0) (1 - x/x_S)^2 \quad (4.2.23)$$

$$\text{If } V_{GB} < V_{GBCrS} \quad f_S(x) = c_0 + c_1x + c_2x^2 \text{ on } [0, x_m] \quad (4.2.24)$$

$$f_S(x) = \psi_{CrS} (1 - x/x_S)^2 \text{ on } [x_m, x_S] \quad (4.2.25)$$

$$\text{If } V_{GB} > V_{GBCrD} \quad f_D(x) = \psi(0,L) (1 - x/x_D)^2 \quad (4.2.26)$$

$$\text{If } V_{GB} < V_{GBCrD} \quad f_D(x) = d_0 + d_1x + d_2x^2 \text{ on } [0, x_m] \quad (4.2.27)$$

$$f_D(x) = \psi_{CrD} (1 - x/x_D)^2 \text{ on } [x_m, x_D] \quad (4.2.28)$$

where V_{GBCrS} , $f_S(x)$, $\psi(0,0)$, x_S , c_0 , c_1 , c_2 and ψ_{CrS} are the source case and V_{GBCrD} , $f_D(x)$, $\psi(0,L)$, x_D , d_0 , d_1 , d_2 and ψ_{CrD} are the drain case of the generic variables V_{GBCr} , $\psi(x)$, ψ_S , x_0 , a_0 , a_1 , a_2 and ψ_{Cr} used in equations (4.2.1) to (4.2.22). The other variables ϕ_n , V_{GBT} , ψ_m and ψ_m must also be set to their appropriate values at the source and drain when $f_S(x)$ and $f_D(x)$ are calculated.

The final requirement to be fulfilled so that (4.2.23), to (4.2.28) can be utilized is to state the methods used to

calculate x_S , x_D and x_m . Since the substrate has been assumed to be uniform x_S and x_D are calculated from (2.1.47) and (2.1.48) whose use is described in Chapter 2. An expression for x_m cannot be derived but must be extracted from MINIMOS simulations. This was done by choosing x_m so that the double parabola approximation for $\psi(x,L)$ fit as closely as possible to the potential profiles extracted from MINIMOS simulations done in saturation. It was found that x_m only varies over a small range and that it is virtually independent of the bias conditions V_{GS} , V_{DS} and V_{SB} , and the oxide thickness t_{ox} . The only two things x_m was found to depend on were the substrate doping N_B and the junction depth r_j . Hence MINIMOS simulations were done to determine the variation of x_m with N_B and r_j . The results are listed in Table 4.1.1.

Table 4.1.1. The matching depth x_m as a function of the substrate doping N_B and junction depth r_j as extracted from MINIMOS simulations.

x_m (μm)	N_B (cm^{-3})		

r_j (μm)	2×10^{15}	2×10^{16}	2×10^{17}
0.2	0.20	0.13	0.10
0.532	0.30	0.25	0.175
1.0	0.50	0.35	0.25

Notice that x_m decreases as the substrate doping increases, and increases as the junction depth increases. This table covers

what is considered to be the entire technologically relevant range but values of x_m outside this range can easily be extracted from MINIMOS simulations. To facilitate easy use of Table 4.1.1 its data was fit with a two-dimensional interpolation polynomial of the form

$$\begin{aligned} x_m = & (b_{00} + b_{01} y + b_{02} y^2) \\ & + (b_{10} + b_{11} y + b_{12} y^2) r_j \\ & + (b_{20} + b_{21} y + b_{22} y^2) r_j^2 \end{aligned} \quad (4.2.29)$$

where the b parameters are given to double precision by

$$b_{00} = 0.1774659689123601 \times 10^{-2} \text{ cm} \quad (4.2.30)$$

$$b_{01} = -.2117465543023180 \times 10^{-3} \text{ cm}$$

$$b_{02} = 0.6325494284831654 \times 10^{-5} \text{ cm}$$

$$b_{10} = -.6961524166444028 \times 10^2$$

$$b_{11} = 0.8552048634168843 \times 10^1$$

$$b_{12} = -.2607796570898981 \times 10^0$$

$$b_{20} = 0.6089936299565081 \times 10^0 \text{ cm}^{-1}$$

$$b_{21} = -.7374634590928879 \times 10^5 \text{ cm}^{-1}$$

$$b_{22} = 0.2225247142415820 \times 10^4 \text{ cm}^{-1}$$

and where $y = \log_{10} N_B$

(4.2.31)

To use (4.2.29) and (4.2.31) N_B must have units of cm^{-3} and r_j must have units of cm to give x_m in cm. With (4.2.29) x_m can be calculated for any values of N_B and r_j in the range $0.2 \mu\text{m} < r_j \leq 1.0 \mu\text{m}$ and $2 \times 10^{15} \text{cm}^{-3} < N_B < 2 \times 10^{17} \text{cm}^{-3}$. Thus the side boundary conditions can now be calculated without the necessity of any further MINIMOS simulations.

To demonstrate the accuracy of the single/double parabola side boundary conditions they are compared to potential profiles extracted from MINIMOS simulations in Figures 4.2.2 and 4.2.3. This was done for a device whose parameters were $L = 1 \mu\text{m}$, $W = 24 \mu\text{m}$, $N_B = 2 \times 10^{16} \text{cm}^{-3}$, $N_S = 2.3 \times 10^{20} \text{cm}^{-3}$, $t_{\text{ox}} = 500 \text{\AA}$, $r_j = 0.532 \mu\text{m}$, and $V_{\text{FB}} = -0.97152 \text{ V}$. The bias conditions used were $V_{\text{GS}} = 1, 3, 5 \text{ V}$, $V_{\text{DS}} = 5 \text{ V}$ and $V_{\text{SB}} = 0, 5 \text{ V}$. The drain voltage was set to 5 V to display the capabilities of the double parabola boundary condition. It is clear from Figures 4.2.2 and 4.2.3 that the single/double parabola side boundary conditions model the true variation of $\psi(x, 0)$ and $\psi(x, L)$ very accurately for this device. Such agreement was typical for the wide range of device structures that were examined. The reason is because the essence of the physical reality has been captured by the formulation of the boundary conditions. The discrepancies that do exist are attributable either to the fact that the true potential variation has a more complicated form than a parabola, or that MINIMOS does not make the sheet charge approximation but the approximate boundary conditions do. The latter reason is the cause of the discrepancy for $V_{\text{GS}} = 3 \text{ V}$ near the surface. But the discrepancies are small and the approximate boundary conditions

(a)

(b)

(c)

Figure 4.2.2. Comparison of the single/double parabola side boundary conditions to the results of MINIMOS simulations where $V_{DS} = 5$ V, $V_{SB} = 0$, (a) $V_{GS} = 1$ V, (b) $V_{GS} = 3$ V and (c) $V_{GS} = 5$ V.

(a)

(b)

(c)

Figure 4.2.3. Comparison of the single/double parabola side boundary conditions to the results of MINIMOS simulations where $V_{DS} = 5$ V, $V_{SB} = 5$ V, (a) $V_{GS} = 1$ V, (b) $V_{GS} = 3$ V and (c) $V_{GS} = 5$ V.

provide a satisfactory description of $\psi(x,0)$ and $\psi(x,L)$ under a wide range of conditions.

4.3 The Long-Channel Trial Function

Now that the side boundary conditions are completely defined and verified the description of the Poisson problem is complete and its solution may begin. The first step is to use the linearity of Poisson's equation to break (4.1.1) into two parts:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = K \quad (4.3.1)$$

$$\frac{\partial^2 R}{\partial x^2} + \frac{\partial^2 R}{\partial y^2} = 0 \quad (4.3.2)$$

$$\text{where } \psi(x,y) = U(x,y) + R(x,y) \quad (4.3.3)$$

and where $U(x,y)$ is the ideal long-channel device trial function and $R(x,y)$ is the solution of Laplace's equation. This approach is similar to that used for the subthreshold solution in that it employs an ideal long-channel trial function to achieve Laplace reduction, and then solves Laplace's equation using the separation of variables technique. The major difference between the two approaches is that in the subthreshold solution the trial function is chosen to be one-dimensional while in the general solution it is chosen to be two-dimensional. It is not absolutely necessary to make the trial function in the general solution two-dimensional. In fact, the same one-dimensional

trial function used in the subthreshold solution could have been employed. But there are two major advantages to using a two-dimensional trial function. First, it allows the trial function $U(x,y)$ to be identified with the potential in an ideal long-channel device which is two-dimensional under general bias conditions. This facilitates an easy physical interpretation of the various parts of the solution that make up the two-dimensional potential $\psi(x,y)$. Second, using the two-dimensional potential for an ideal long-channel device allows (4.1.5) to be used as the boundary condition on the bottom of the rectangular region. If a one-dimensional trial function was used, then the derivative $\partial\psi/\partial x(d,y)$ in (4.1.5) would have to be set to zero, making this boundary condition less accurate especially above threshold.

The disadvantage to using a two-dimensional trial function is that the gradual channel approximation which states that

$$\frac{\partial^2 U}{\partial x^2} \gg \frac{\partial^2 U}{\partial y^2} \quad (4.3.4)$$

must be employed to reduce (4.3.1) to

$$\frac{\partial^2 U(x,y)}{\partial x^2} = K \quad (4.3.5)$$

so that $U(x,y)$ may be solved for analytically. But fortunately the error incurred by making the gradual channel approximation for $U(x,y)$ is almost negligible even for devices with channel

lengths as small as $0.5 \mu\text{m}$. This is demonstrated by plotting $(\partial^2 U / \partial y^2) / K$ as a percentage error along the silicon surface for two devices with channel lengths of $1 \mu\text{m}$ and $0.5 \mu\text{m}$ respectively, and whose other parameters are given in Table 4.3.1. The closed form expression for $\partial^2 U / \partial y^2$ is given by equation (4.3.47) and is derived in due course.

Table 4.3.1. Parameters for Devices 1 and 2.

Device Type	L	W	N_B	N_S	t_{ox}	r_j	V_{FB}
	(μm)	(μm)	(cm^{-3})	(cm^{-3})	($\text{\AA}$)	(μm)	(V)
Device 1	1.0	24	2×10^{16}	2.3×10^{20}	500	0.532	-0.72
Device 2	0.5	24	4×10^{16}	2.0×10^{20}	150	0.2	-0.207

The percentage error curves and surface potentials for Devices 1 and 2 are displayed in Figures 4.3.1 and 4.3.2 respectively. The bias conditions used for both these figures were $V_{GS} = 5 \text{ V}$, $V_{DS} = 5 \text{ V}$ and $V_{SB} = 0$ in order to test the gradual channel approximation under severe conditions. These two figures show that the error in the gradual channel approximation is generally less than $10^{-3}\%$ throughout the channel. Only at the drain boundary does the error even start to become significant, as at this point it is 0.86% for Device 1 and 4.82% for Device 2. Thus even for devices with channel lengths as short as $0.5 \mu\text{m}$ the gradual channel approximation is an excellent approximation to

(a)

(b)

Figure 4.3.1. (a) Percentage error in the gradual channel approximation versus position along the channel y , and (b) the long-channel surface potential, $U(0,y)$ versus position along the channel y , for Device 1 where $V_{GS} = 5$ V, $V_{DS} = 5$ V and $V_{SB} = 0$.

(a)

(b)

Figure 4.3.2. (a) Percentage error in the gradual channel approximation versus position along the channel y , and (b) the long-channel surface potential $U(0,y)$ versus position along the channel y , for Device 2 where $V_{GS} = 5 \text{ V}$, $V_{DS} = 5 \text{ V}$ and $V_{SB} = 0$.

make in solving for $U(x,y)$. Hence the ideal long-channel potential $U(x,y)$ is internally consistent with respect to the gradual channel approximation, because assuming the approximation results in an answer that agrees with the approximation. It must be noted, however, that the gradual channel approximation is not valid for the real two-dimensional potential $\psi(x,y)$, because if it was then short-channel modelling could be achieved by solving only a one-dimensional Poisson's equation, which is not the case.

Solving (4.3.5) and using the two depletion edge boundary conditions

$$U(x_d(y), y) = 0 \quad (4.3.6)$$

$$\frac{\partial U}{\partial x}(x_d(y), y) = 0 \quad (4.3.7)$$

yields the form of $U(x,y)$:

$$U(x,y) = \frac{K}{2} (x_d - x)^2 \quad (4.3.8)$$

which by noting that

$$U(0,y) = \frac{K}{2} x_d^2(y) \quad (4.3.9)$$

may be converted to

$$U(x,y) = U(0,y) (1 - x/x_d)^2 \quad (4.3.10)$$

Since (4.3.9) can be arranged to

$$x_d(y) = \sqrt{\frac{2}{K} U(0,y)} \quad (4.3.11)$$

it is clear that solving for the long-channel surface potential $U(0,y)$ yields $x_d(y)$ from which $U(x,y)$ can be obtained using (4.3.8) or (4.3.10).

The solution for $U(0,y)$ starts with the one-dimensional current transport equation

$$I_D = \mu_n W \left(-Q_m \frac{dU_s}{dy} + \frac{1}{\beta} \frac{dQ_m}{dy} \right) \quad (4.3.12)$$

$$\text{where } U_s(y) = U(0,y) \quad (4.3.13)$$

Current continuity is then invoked by rearranging (4.3.12) and setting the constant P equal to

$$P = \frac{I_D}{\mu_n W} = -Q_m \frac{dU_s}{dy} + \frac{1}{\beta} \frac{dQ_m}{dy} \quad (4.3.14)$$

which assumes that the mobility is constant. It must be made clear that assuming a constant mobility in the derivation of the trial function $U(\bar{x},y)$ does not preclude using a field dependent mobility in the overall model. This is because the purpose of the trial function is simply the mathematical reduction of Poisson's equation to Laplace's equation, which has no bearing on the physics of the problem. Thus it is still permissible to include a field dependent mobility in the current continuity equation and hence in the solution for $\psi(x,y)$.

Integrating (4.3.14) gives

$$P \cdot y + Q = - \int Q_m dU_s + \frac{Q_m}{\beta} \quad (4.3.15)$$

Now from (4.3.10) and (4.3.11) it can be derived that

$$\frac{\partial U}{\partial x}(0, y) = - \sqrt{\frac{2q N_B}{\epsilon_s}} U_S \quad (4.3.16)$$

Substituting (4.3.16) into the statement of the continuity of normal displacement for a sheet channel which is

$$Q_m = - C_{ox} (V_{GB}' - U_S) - \epsilon_s \frac{\partial U}{\partial x}(0, y) \quad (4.3.17)$$

gives

$$Q_m = - C_{ox} (V_{GB}' - U_S) + \sqrt{2 \epsilon_s q N_B} U_S \quad (4.3.18)$$

Substituting (4.3.18) into (4.3.15), integrating and rearranging yields

$$\frac{C_{ox}}{2} (V_{GB}' - U_S)^2 + \frac{2}{3} \sqrt{2 \epsilon_s q N_B} U_S^{3/2} + \frac{C_{ox}}{\beta} (V_{GB}' - U_S) - \frac{1}{\beta} \sqrt{2 \epsilon_s q N_B} U_S^{1/2} + P y + Q = 0 \quad (4.3.19)$$

Rearranging yet again to

$$U_S^2 + \frac{4}{3 C_{ox}} \sqrt{2 \epsilon_s q N_B} U_S^{3/2} - 2 (V_{GB}' + \frac{1}{\beta}) U_S - \frac{2}{\beta C_{ox}} \sqrt{2 \epsilon_s q N_B} U_S^{1/2} + V_{GB}' (V_{GB}' + 2/\beta) + \frac{2}{C_{ox}} (P y + Q) = 0 \quad (4.3.20)$$

3

yields a fourth order polynomial in $U_s^{1/2}$. This is the highest order polynomial whose roots have a closed form solution. Thus solving for the correct one of the four roots and squaring yields a closed form solution for U_s :

$$U_s = \left[-\frac{a}{4} + \frac{R_q - \sqrt{\frac{3a^2}{4} - R_q^2 - 2b + \frac{4ab - 8c - a^3}{4 R_q}}}{2} \right]^2 \quad (4.3.21)$$

where

$$R_q = \sqrt{\frac{a^2}{4} - \frac{2b}{3}} + 2\sqrt{\frac{a}{3}} \cos\left(\frac{\cos^{-1} \sqrt{[2b^3 - 9b(ac + 8d) + 27(a^2d + c^2)]^2}}{108 a_c^3}}\right) \quad (4.3.22)$$

$$a_c = \frac{b^2}{3} - ac + 4d \quad (4.3.23)$$

$$a = \frac{4}{3 C_{ox}} \sqrt{2 \epsilon_s q N_B} \quad (4.3.24)$$

$$b = -2 (V_{GB}' + 1/\beta) \quad (4.3.25)$$

$$c = \frac{-2}{\beta C_{ox}} \sqrt{2 \epsilon_s q N_B} \quad (4.3.26)$$

$$d = V_{GB}' (V_{GB}' + 2/\beta) + \frac{2}{C_{ox}} (P y + Q) \quad (4.3.27)$$

The constants P and Q are determined using the boundary potentials $U(0,0)$ and $U(0,L)$ by evaluating (4.3.19) at $y = 0$ and $y = L$. Solving the resulting two equations for P and Q yields

$$Q = -C_{ox} (V_{GB} - U_{SS}) \left(\frac{V_{GB} - U_{SS}}{2} + \frac{1}{\beta} \right) - \sqrt{2 \epsilon_s q N_B U_{SS}} \left(\frac{2}{3} U_{SS} - \frac{1}{\beta} \right) \quad (4.3.28)$$

$$R_B = -C_{ox} (V_{GB} - U_{SD}) \left(\frac{V_{GB} - U_{SD}}{2} + \frac{1}{\beta} \right) - \sqrt{2 \epsilon_s q N_B U_{SD}} \left(\frac{2}{3} U_{SD} - \frac{1}{\beta} \right) \quad (4.3.29)$$

$$P = \frac{R_B - Q}{L} \quad (4.3.30)$$

$$\text{where } U_{SS} = U(0,0) \quad (4.3.31)$$

$$U_{SD} = U(0,L) \quad (4.3.32)$$

The boundary potentials $U(0,0)$ and $U(0,L)$ are calculated from the solution of a one-dimensional Shockley-Poisson equation along the lines $y = 0$ and $y = L$ by applying the principle of the continuity of normal displacement at the silicon surface in the manner originated by Pao and Sah [2]. This derivation follows a pattern similar to that required for the calculation of the surface potentials $\psi(0,0)$ and $\psi(0,L)$ needed by the single parabola side boundary conditions, so its treatment is brief. Note that the potentials $U(0,0)$ and $U(0,L)$ are different from the potentials $\psi(0,0)$ and $\psi(0,L)$ because the former potentials are the boundary potentials for an ideal long-channel device, while the latter potentials are the true electrostatic potentials at points $(0,0)$ and $(0,L)$.

The one-dimensional Shockley-Poisson equation for $U(x,y)$ is

$$\frac{d^2U}{dx^2} = \frac{1}{2} \frac{dE_x^2}{dU} = \frac{q}{\epsilon_s} [n_i e^{\beta(U-\phi_n)} - n_i e^{\beta(\phi_p-U)} + N_B] \quad (4.3.33)$$

Integrating from the depletion edge to the surface gives the square of the normal surface field as

$$E_s^2 = \frac{2q}{\beta\epsilon_s} [n_i e^{\beta(U_s-\phi_n)} - n_i e^{-\beta\phi_n} + n_i e^{\beta(\phi_p-U_s)} - n_i e^{\beta(\phi_p-U_s)} + N_B U_s] \quad (4.3.34)$$

Neglecting the insignificant term $n_i e^{-\beta\phi_n}$ and the term $n_i e^{\beta(\phi_p-U_s)}$, which is significant only in accumulation, and using the fact that

$$E_s = \frac{C_{ox}}{\epsilon_s} (V_{GB} - U_s) \quad (4.3.35)$$

derived from a one-dimensional solution of Laplace's equation in the gate oxide, then (4.3.34) becomes

$$n_i e^{\beta(U_s-\phi_n)} + N_B (\beta U_s - 1) - \beta C (V_{GB} - U_s)^2 = 0 \quad (4.3.36)$$

$$\text{where } C = \frac{C_{ox}^2}{2q\epsilon_s} \quad (4.3.37)$$

The differentiation of (4.3.36) permits the construction of the Newton-Raphson iterative equation that is used to calculate U_S .

Thus

$$U'_S = U_S - \frac{1}{\beta} \frac{n_i e^{\beta(U_S - \phi_n)} + N_B (\beta U_S - 1) - \beta C (V_{GB} - U_S)^2}{n_i e^{\beta(U_S - \phi_n)} + N_B + 2C (V_{GB} - U_S)} \quad (4.3.38)$$

where U'_S is the next approximation to the true value of U_S in the iterative sequence.

To make efficient use of (4.3.38) it is necessary to obtain two good initial guesses for U_S : one that is valid above threshold and the other that is valid below. The threshold voltage that defines the boundary between these two regimes is found by substituting the threshold condition $U_S = \phi$ into (4.3.36), neglecting the electron charge term and rearranging. This gives

$$V_{GBT} = \phi + \sqrt{\frac{N_B}{C}} \phi \quad (4.3.39)$$

$$\text{where } \phi = \phi_n + \phi_p \quad (4.3.40)$$

$$\text{and } \phi_n(0) = \phi_p + V_{SB} \quad (4.3.41)$$

$$\phi_n(L) = \phi_p + V_{SB} + V_{DS} \quad (4.3.42)$$

The initial guess for U_S above threshold is obtained by starting with the preliminary guess of $U_S = \phi$, substituting this into

(4.3.36) everywhere except in the exponential, and then isolating the U_s in the exponential. This gives

$$U_s = \phi + \frac{1}{\beta} \ln \left[\frac{\beta C (V_{GB} - \phi)^2 - N_B (\beta \phi - 1)}{n_i} \right] \quad (4.3.43)$$

To obtain the below threshold initial guess neglect the electron charge term in (4.3.36) and solve the resulting quadratic equation. This yields

$$U_s = V_{GB} + A \left(1 - \sqrt{1 + \frac{2}{A} (V_{GB} - 1/\beta)} \right) \quad (4.3.44)$$

$$\text{where } A = \frac{N_B}{2C} \quad (4.3.45)$$

The procedure for calculating U_s is similar to that used for calculating the surface potential ψ_s required by the single parabola boundary condition. It starts with the appropriate value for ϕ_n given by (4.3.41) for $U(0,0)$ and (4.3.42) for $U(0,L)$. Then the threshold voltage is calculated from (4.3.39), and (4.3.43) is used as an initial guess if $V_{GB} > V_{GBT}$, or (4.3.44) is used as an initial guess if $V_{GB} < V_{GBT}$. Iterating using (4.3.38) a maximum of two times determines U_s to seven significant digits. Although the procedures for calculating ψ_s and U_s are similar the equations used are quite different. This is because (4.2.9) makes the sheet charge approximation while (4.3.36) does not. Furthermore, (4.2.9) obtains the electric field below the silicon surface from a parabolic approximation to

the potential while (4.3.36) obtains it from the solution of the Shockley-Poisson equation.

To complete the discussion of the trial function $U(x,y)$ it is necessary to derive an expression for the second derivative $\partial^2 U / \partial y^2(0,y)$ because it was used in Figures 4.3.1 and 4.3.2 to prove the validity of employing the gradual channel approximation in the calculation of $U(x,y)$. The first derivative is obtained by differentiating (4.3.19) to give

$$\frac{\partial U}{\partial y}(0,y) = \frac{P}{C_{ox}(V_{GB} - U_S + 1/\beta) - \sqrt{2\epsilon_s q N_B U_S (1 - \frac{1}{2\beta U_S})}} \quad (4.3.46)$$

Differentiating again and rearranging produces the second derivative.

$$\frac{\partial^2 U}{\partial y^2}(0,y) = \frac{P^2}{D^3} [C_{ox} + \sqrt{\frac{\epsilon_s q N_B}{2 U_S}} (1 + \frac{1}{2\beta U_S})] \quad (4.3.47)$$

$$\text{where } D = C_{ox}(V_{GB} - U_S + 1/\beta) - \sqrt{2\epsilon_s q N_B U_S (1 - \frac{1}{2\beta U_S})} \quad (4.3.48)$$

The error function that was plotted in Figures 4.3.1 and 4.3.2 was calculated by taking the result of (4.3.47) and dividing it by K .

4.4 Solution of the Laplace Problem

The description of the long-channel trial function $U(x,y)$ is now complete and the Laplace problem that results from the use of $U(x,y)$ may be described. The boundary conditions for the Laplace

problem are obtained from the boundary conditions for the Poisson problem by rearranging (4.3.3) to

$$R(x,y) = \psi(x,y) - U(x,y) \quad (4.4.1)$$

The boundary conditions for the sides and bottom of the rectangular region then follow immediately. To determine the surface boundary condition the mobile charge density function $Q_m(y)$ must be divided into a long-channel part $Q_{mLC}(y)$ that is associated with $U(x,y)$, and a short-channel part $Q_{mSC}(y)$ that is associated with $R(x,y)$. Hence

$$Q_m(y) = Q_{mLC}(y) + Q_{mSC}(y) \quad (4.4.2)$$

This then permits (4.1.4) to be divided as follows

$$\begin{aligned} & \{ C_{ox} [V_{GB} - U(0,y)] + Q_{mLC}(y) + \epsilon_s \frac{\partial U}{\partial x}(0,y) \} \\ & + \{ \epsilon_s \frac{\partial R}{\partial x}(0,y) - C_{ox} R(0,y) + Q_{mSC}(y) \} = 0 \end{aligned} \quad (4.4.3)$$

The expression in the first brace bracket of (4.4.3) is a statement of the continuity of normal displacement at the silicon-silicon dioxide interface for the long-channel trial function $U(x,y)$. Hence the value of this brace bracket is zero. Upon defining for convenience the short-channel voltage function

$$V_{SC}(y) = - \frac{Q_{mSC}(y)}{C_{ox}} \quad (4.4.4)$$

the Laplace problem can be written down by recalling (4.3.2), using (4.4.1) to obtain the side and bottom boundary conditions and (4.4.3) to obtain the top boundary condition. Thus the Laplace problem is

$$\frac{\partial^2 R}{\partial x^2} + \frac{\partial^2 R}{\partial y^2} = 0 \quad (4.4.5)$$

$$R(x, 0) = f_S(x) - \frac{K}{2} [x_d(0) - x]^2 \quad (4.4.6)$$

$$R(x, L) = f_D(x) - \frac{K}{2} [x_d(L) - x]^2 \quad (4.4.7)$$

$$R \tau_{ox} \frac{\partial R}{\partial x} (0, y) - R(0, y) = V_{SC}(y) \quad (4.4.8)$$

$$\frac{\partial R}{\partial x} (d, y) = 0 \quad (4.4.9)$$

where $f_S(x)$ and $f_D(x)$ are given by (4.2.23) to (4.2.28) and $x_d(0)$ and $x_d(L)$ are defined by (4.3.11).

This Laplace problem has three nonhomogeneous boundary conditions namely (4.4.6), (4.4.7) and (4.4.8). They are nonhomogeneous because they involve functions other than $R(x, y)$ and its first derivatives. The most that can be handled in a single separation of variables solution is two nonhomogeneous boundary conditions on opposite sides of the rectangular region under consideration. This is because the last step in a separation of variables solution is a Fourier analysis of the nonhomogeneous function, and in order to carry this out the sin and/or cos eigenfunctions must be oriented in the same direction

(e.g. the x direction) as the nonhomogeneous function. The $\sinh$ and/or $\cosh$ eigenfunctions cannot be oriented in this direction because simple Fourier analysis cannot be done with them as they are nonorthogonal eigenfunctions. Hence to effect a solution of the Laplace problem described by (4.4.5) to (4.4.9), the linearity of Laplace's equation must be used to divide this problem into two new Laplace problems. This is done by letting

$$R(x,y) = V(x,y) + Z(x,y) \quad (4.4.10)$$

thereby creating the two new Laplace problems

Laplace Problem 1:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad (4.4.11)$$

$$V(x,0) = f_S(x) - \frac{K}{2} [x_d(0) - x]^2 \quad (4.4.12)$$

$$V(x,L) = f_D(x) - \frac{K}{2} [x_d(L) - x]^2 \quad (4.4.13)$$

$$R \left. \frac{\partial V}{\partial x} \right|_{x=0} (0,y) - V(0,y) = 0 \quad (4.4.14)$$

$$\frac{\partial V}{\partial x} (d,y) = 0 \quad (4.4.15)$$

Laplace Problem 2:

$$\frac{\partial^2 Z}{\partial x^2} + \frac{\partial^2 Z}{\partial y^2} = 0 \quad (4.4.16)$$

$$Z(x,0) = 0 \quad (4.4.17)$$

$$Z(x, L) = 0 \quad (4.4.18)$$

$$R \lim_{x \rightarrow 0} \frac{\partial Z}{\partial x}(0, y) - Z(0, y) = V_{SC}(y) \quad (4.4.19)$$

$$\frac{\partial Z}{\partial x}(d, y) = 0 \quad (4.4.20)$$

Note that the boundary conditions of these two Laplace problems sum to give the boundary conditions of the original Laplace problem. Laplace problem 1 can be solved in a manner that orients the sin and/or cos eigenfunctions in the x direction and Laplace problem 2 can be solved in a manner that orients the sin and/or cos eigenfunctions in the y direction. This will permit simple Fourier analysis of the nonhomogeneous functions.

The first Laplace problem is of the same form as the Laplace problem described by (2.1.4c), (2.2.13) and (2.2.17) to (2.2.19) in Chapter 2. Hence its solution is of the same form which is repeated here for convenience.

$$V(x, y) = \sum_{i=0}^{\infty} \frac{\cos[u_i(1-x/d)]}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \quad (2.2.21)$$

$$\text{where } A_i = \frac{1}{d} \int_0^d V(x, 0) \cos[u_i(1-x/d)] dx \quad (2.2.22)$$

$$B_i = \frac{1}{d} \int_0^d V(x, L) \cos[u_i(1-x/d)] dx \quad (2.2.23)$$

$$\alpha_i = \frac{1}{2} \left[1 + \frac{\sin(2 u_i)}{2 u_i} \right] \quad (2.2.24)$$

and where the eigenvalues u_i are defined by

$$\tan u_i = R \frac{d}{t_{ox}} u_i \quad (2.2.25)$$

The method used to calculate the eigenvalues u_i is exactly the same as described in Chapter 2.

The difference between the solution given for $V(x,y)$ in Chapter 2 and the solution that is presently sought, is that here the single/double parabola boundary conditions change the form of $V(x,0)$ and $V(x,L)$. This produces a major difference in the nature of the Fourier coefficients A_i and B_i because their formulation will depend upon whether a single or double parabola boundary condition is used and whether d is greater or less than x_m . For the single parabola boundary condition, which is the case when $V_{GB}' > V_{GBCrS}'$ or $V_{GB}' > V_{GBCrD}'$, A_i and B_i are calculated from (2.2.22) and (2.2.23) using (4.4.12), (4.4.13), (4.2.23) and (4.2.26). The solutions of the three integrals that must be done in this calculation are given by (2.2.37), (2.2.38) and (2.2.39). The results are

$$\begin{aligned}
 A_i = & (\psi_{SS} - U_{SS}) \frac{\sin u_i}{u_i} + \frac{d}{x_S} \left(Kd - \frac{2 \psi_{SS}}{x_S} \right) \frac{1 - \cos u_i}{u_i^2} \\
 & + 2d^2 \left(\frac{\psi_{SS}}{x_S^2} - M \right) \frac{u_i - \sin u_i}{u_i^3}
 \end{aligned} \quad (4.4.21)$$

$$\begin{aligned}
 B_i = & (\psi_{SD} - U_{SD}) \frac{\sin u_i}{u_i} + d \left[K x_d(L) - \frac{2 \psi_{SD}}{x_D} \right] \frac{1 - \cos u_i}{u_i^2} \\
 & + 2d^2 \left(\frac{\psi_{SD}}{x_D^2} - M \right) \frac{u_i - \sin u_i}{u_i^3}
 \end{aligned} \quad (4.4.22)$$

Note that (4.4.21) appears to be slightly different in form than (4.2.22) because $x_d(0)$ is replaced by d because the two quantities are equivalent. The fact that $x_d(0) = d$ is a consequence of defining d to be the minimum depletion depth for the ideal long-channel device (see equations (4.1.8) and (4.3.11)).

If $V_{GB} \leq V_{GBCrS}$ or $V_{GB} \leq V_{GBCrD}$ then it is necessary to use the double parabola boundary conditions. The calculations for A_i and B_i can then take on two different forms. The simpler case occurs when $d \leq x_m$ then the integrals in (2.2.22) and (2.2.23) can be done using (2.2.37), (2.2.38) and (2.2.39). But for the more complex case, that occurs when $d > x_m$, the integrals in (2.2.22) and (2.2.23) must be divided into an integration over the interval $[0, x_m]$ and an integration over the interval $[x_m, d]$, since the integrand is different on these two intervals due to

the use of the two parabolas in the double parabola boundary condition. The integrals that are needed to perform the integrations when $d > x_m$ are

$$\frac{1}{d} \int_0^{x_m} \cos[u_i(1 - x/d)] dx = \frac{1}{u_i} \{ \sin u_i - \sin[u_i(1 - x_m/d)] \} \quad (4.4.23)$$

$$\begin{aligned} \frac{1}{d} \int_0^{x_m} x \cos[u_i(1 - x/d)] dx &= \frac{d}{u_i^2} \{ \cos[u_i(1 - x_m/d)] - \cos u_i \} \\ &- \frac{x_m}{u_i} \sin[u_i(1 - x_m/d)] \end{aligned} \quad (4.4.24)$$

$$\begin{aligned} \frac{1}{d} \int_0^{x_m} x^2 \cos[u_i(1 - x/d)] dx &= \frac{2d}{u_i^3} x_m \cos[u_i(1 - x_m/d)] \\ &- \frac{d^2}{u_i^3} \{ 2 \sin u_i + (u_i^2 x_m^2 / d^2 - 2) \sin[u_i(1 - x_m/d)] \} \end{aligned} \quad (4.4.25)$$

$$\frac{1}{d} \int_{x_m}^d \cos[u_i(1 - x/d)] dx = \frac{\sin[u_i(1 - x_m/d)]}{u_i} \quad (4.4.26)$$

$$\begin{aligned} \frac{1}{d} \int_{x_m}^d x \cos[u_i(1 - x/d)] dx &= \frac{d}{u_i^2} \{ 1 - \cos[u_i(1 - x_m/d)] \} \\ &+ \frac{x_m}{u_i} \sin[u_i(1 - x_m/d)] \end{aligned} \quad (4.4.27)$$

$$\frac{1}{d} \int_{x_m}^d x^2 \cos \{u_i (1 - x/d)\} dx = \frac{2d}{u_i^2} (d - x_m \cos[u_i (1 - x_m/d)]) + \frac{d^2}{u_i^3} (u_i^2 x_m^2/d^2 - 2) \sin[u_i (1 - x_m/d)] \quad (4.4.28)$$

For the simpler case where $d < x_m$, A_i and B_i are calculated from (2.2.22) and (2.2.23) using (4.4.12), (4.4.13), (4.2.24) and (4.2.27). Performing the integrals using (2.2.37), (2.2.38) and (2.2.39) yields

$$A_i = (c_0 - U_{SS}) \frac{\sin u_i}{u_i} + d (c_1 + Kd) \frac{1 - \cos u_i}{u_i^2} + 2d^2 (c_2 - M) \frac{u_i - \sin u_i}{u_i^3} \quad (4.4.29)$$

$$B_i = (d_0 - U_{SD}) \frac{\sin u_i}{u_i} + d [d_1 + K x_d(L)] \frac{1 - \cos u_i}{u_i^2} + 2d^2 (d_2 - M) \frac{u_i - \sin u_i}{u_i^3} \quad (4.4.30)$$

For the case where $d > x_m$, A_i and B_i are obtained from (2.2.22) and (2.2.23) using (4.4.12), (4.4.13), (4.2.24), (4.2.25), (4.2.27) and (4.2.28). The integrals are performed using (4.4.23) to (4.4.28) and then the following expressions are obtained.

$$\begin{aligned}
A_i = & (c_0 - U_{SS}) \frac{1}{u_i} \{ \sin u_i - \sin [u_i (1 - x_m/d)] \} \\
& + (c_1 + Kd) \left\{ \frac{d}{u_i^2} [\cos[u_i (1 - x_m/d)] - \cos u_i] - \frac{x_m}{u_i} \sin[u_i (1 - x_m/d)] \right\} \\
& + (c_2 - M) \left\{ \frac{2d}{u_i^2} x_m \cos [u_i (1 - x_m/d)] \right. \\
& \left. - \frac{d^2}{u_i^3} [2 \sin u_i + (u_i^2 x_m^2 / d^2 - 2) \sin[u_i (1 - x_m/d)]] \right\} \\
& + (\psi_{CrS} - U_{SS}) \frac{\sin [u_i (1 - x_m/d)]}{u_i} \\
& + (Kd - \frac{2 \psi_{CrS}}{x_S}) \left\{ \frac{d}{u_i^2} [1 - \cos[u_i (1 - x_m/d)]] + \frac{x_m}{u_i} \sin[u_i (1 - x_m/d)] \right\} \\
& + \left(\frac{\psi_{CrS}}{x_S^2} - M \right) \left\{ \frac{2d}{u_i^2} [d - x_m \cos [u_i (1 - x_m/d)]] \right. \\
& \left. + \frac{d^2}{u_i^3} (u_i^2 x_m^2 / d^2 - 2) \sin [u_i (1 - x_m/d)] \right\} \quad (4.4.31)
\end{aligned}$$

$$\begin{aligned}
B_i = & (d_0 - U_{SD}) \frac{1}{u_i} \{ \sin u_i - \sin [u_i (1 - x_m/d)] \} \\
& + \{ d_1 + Kx_d(L) \} \left\{ \frac{d}{u_i^2} [\cos [u_i (1 - x_m/d)] - \cos u_i] - \frac{x_m}{u_i} \sin [u_i (1 - x_m/d)] \right\} \\
& + (d_2 - M) \left\{ \frac{2d}{u_i^2} x_m \cos [u_i (1 - x_m/d)] \right. \\
& \left. - \frac{d^2}{u_i^3} [2 \sin u_i + (u_i^2 x_m^2 / d^2 - 2) \sin [u_i (1 - x_m/d)]] \right\} \\
& + (\psi_{CrD} - U_{SD}) \frac{\sin [u_i (1 - x_m/d)]}{u_i} \\
& + [Kx_d(L) - \frac{2}{x_D} \psi_{CrD}] \left\{ \frac{d}{u_i^2} [1 - \cos [u_i (1 - x_m/d)]] + \frac{x_m}{u_i} \sin [u_i (1 - x_m/d)] \right\} \\
& + (\frac{\psi_{CrD}}{x_D^2} - M) \left\{ \frac{2d}{u_i^2} [d - x_m \cos [u_i (1 - x_m/d)]] \right. \\
& \left. + \frac{d^2}{u_i^3} (u_i^2 x_m^2 / d^2 - 2) \sin [u_i (1 - x_m/d)] \right\} \quad (4.4.32)
\end{aligned}$$

With the calculation of the coefficients A_i and B_i the first Laplace problem is solved and the second Laplace problem can now be dealt with. This Laplace problem is solved using the separation of variables technique which assumes a solution of the form

$$Z(x,y) = \sum_{n=0}^{\infty} X_n(x) Y_n(y) \quad (4.4.33)$$

Substituting (4.4.33) into Laplace's equation and applying the boundary conditions (4.4.17), (4.4.18) and (4.4.20) yields

$$Z(x,y) = \sum_{n=1}^{\infty} E_n \frac{\cosh\left[\frac{n\pi}{L}(d-x)\right]}{\cosh\left(\frac{n\pi d}{L}\right)} \sin\left(\frac{n\pi y}{L}\right) \quad (4.4.34)$$

It should be noted that in this Laplace solution the eigenvalues take the simple form of $n\pi/L$. Substituting (4.4.34) into the surface boundary condition (4.4.19) and rearranging gives

$$-\sum_{n=1}^{\infty} E_n \left[1 + R t_{ox} \frac{n\pi}{L} \tanh\left(\frac{n\pi d}{L}\right)\right] \sin\left(\frac{n\pi y}{L}\right) = V_{SC}(y) \quad (4.4.35)$$

Multiplying each side by $\sin(k\pi y/L)$, integrating from 0 to L , and interchanging the order of summation and integration yields

$$\begin{aligned} & -\sum_{n=1}^{\infty} E_n \left[1 + R t_{ox} \frac{n\pi}{L} \tanh\left(\frac{n\pi d}{L}\right)\right] \int_0^L \sin\left(\frac{n\pi y}{L}\right) \sin\left(\frac{k\pi y}{L}\right) dy \\ & = \int_0^L V_{SC}(y) \sin\left(\frac{k\pi y}{L}\right) dy \end{aligned} \quad (4.4.36)$$

The orthogonality relation

$$\int_0^L \sin\left(\frac{n\pi y}{L}\right) \sin\left(\frac{k\pi y}{L}\right) dy = \frac{L}{2} \delta_{nk} \quad (4.4.37)$$

can then be taken advantage of to eliminate the summation and reduce (4.4.36) to

$$E_n = -\frac{2}{L} \frac{\int_0^L V_{SC}(y) \sin\left(\frac{n\pi y}{L}\right) dy}{1 + R t_{ox} \frac{n\pi}{L} \tanh\left(\frac{n\pi d}{L}\right)} \quad (4.4.38)$$

which defines the Fourier coefficients E_n .

The apparent simplicity of (4.4.38) is deceptive as the solution of this equation is the subject of the next two chapters. Its inherent complexity can be appreciated by writing down the complete solution for the two-dimensional potential $\psi(x,y)$. This is obtained by substituting (2.2.21) and (4.4.34) into (4.4.10), and substituting (4.4.10) and (4.3.8) into (4.3.3). This yields

$$\begin{aligned} \psi(x,y) = & \frac{K}{2} [x_d(y) - x]^2 \\ & + \sum_{i=0}^{\infty} \frac{\cos[u_i(1-x/d)]}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \\ & + \sum_{n=1}^{\infty} E_n \frac{\cosh[\frac{n\pi}{L}(d-x)]}{\cosh(\frac{n\pi d}{L})} \sin\left(\frac{n\pi y}{L}\right) \end{aligned} \quad (4.4.39)$$

Everything in (4.4.39) has been determined except for the Fourier coefficients E_n . These are obtained from (4.4.38) by Fourier

analysis of $V_{SC}(y)$, where $V_{SC}(y)$ is a linear function of $Q_m(y)$. But $Q_m(y)$ is calculated from $\psi(0,y)$ and $\phi_n(y)$ using (4.2.3). It is shown in Chapter 5 that $\phi_n(y)$ can be calculated from $\psi(0,y)$ by applying the principle of current continuity. So in essence $Q_m(y)$ has an exponential dependence on $\psi(0,y)$ which depends linearly on the E_n , but a Fourier analysis of $Q_m(y)$ (in the guise of $V_{SC}(y)$) is necessary to calculate the E_n using (4.4.38). Thus a transcendental relationship between a set of highly nonlinear functions is what defines the Fourier coefficients E_n . This is a consequence of the nonlinear coupling between Poisson's equation and the current continuity equation which is expressed by the Poisson problem stated by equations (4.1.1) to (4.1.7).

While casting the remainder of the solution in terms of the E_n presents a major computational problem, it also creates a major opportunity to obtain an elegant solution, because used wisely the E_n can be a powerful concept. This power stems from the amount of information that is represented by the E_n , as choosing the correct E_n values satisfies the principles of both current continuity and the continuity of normal displacement, and provides the two-dimensional potential $\psi(x,y)$ by employing (4.4.39). Thus the entire problem of satisfying Poisson's equation, the current continuity equation, and the boundary conditions has now been reduced to simply solving for the E_n . In practice, since the Fourier series must be truncated for practical computations, there are only 40 or 50 E_n to be determined. Thus the only numerical requirements for finding a solution are to obtain these 40 or 50 numbers. Upon comparing

this to the thousands of similar values that must be calculated in solving Poisson's equation and the current continuity equation by the numerical methods of finite elements or finite differences, the elegance of the analytical approach becomes apparent.

There are two ways to proceed with the calculation of the E_n values. The first is to deal directly with the nonlinear functions just mentioned and iteratively choose the E_n so that the properties of: satisfaction of Poisson's equation, current continuity and continuity of normal displacement are met. While this is the most immediately obvious approach, it is a poor technique because it is computationally very slow and suffers from noise problems. However, this method of solution is explained in Chapter 5 to provide insight into the problem. The better method of calculating the E_n is to transform the whole problem into the Fourier domain where the nonlinear functions become a set of nonlinear algebraic equations that can be solved using elegant techniques. This method is presented in Chapter 6. In some respects this is the more obvious method to try because it is the standard way to finish a separation of variables solution of Laplace's equation. But the algebraic manipulations necessary to achieve it are not at all obvious due to the complexity and nonlinearity of the surface boundary condition (4.1.4).

CHAPTER 5

THE ITERATIVE FUNCTION TECHNIQUE

5.1 Outline of the Technique

The form of the two-dimensional potential $\psi(x,y)$ that satisfies Poisson's equation and the side and bottom boundary conditions (4.1.2), (4.1.3) and (4.1.5) for the rectangular region under consideration is given by (4.4.39). The only quantities that are as yet undetermined in (4.4.39) are the Fourier coefficients E_n . These are chosen so that the surface boundary condition (4.1.4), which entails the continuity of both normal displacement and current, is satisfied. Upon making the sheet charge approximation, so that the inversion layer is modelled as an infinitely thin sheet of charge along the silicon surface, the surface boundary condition is reduced to one dimension. The problem then involves calculating the surface potential $\psi(0,y)$, the quasi-Fermi potential $\phi_n(y)$, and the surface field below the inversion layer $\partial\psi/\partial x(0,y)$ so that they satisfy Poisson's equation, current continuity and the continuity of normal displacement. Mathematically this is equivalent to solving a system of three coupled nonlinear ordinary differential equations in three unknown functions. This is not a simple problem.

The solution of this problem can be obtained by using an iterative function technique. This involves casting each of the three required conditions in a form where if two of the three

functions are known then the third can be calculated. Once initial guesses for the three functions $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$ are chosen, an iterative sequence can then be initiated to cycle through each of the three requirements: satisfaction of Poisson's equation; continuity of normal displacement; and current continuity, until the iterative process converges.

One of the three requirements has already been cast in a suitable form because Poisson's equation has been solved to yield (4.4.39). From this equation the formulations of $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ that satisfy Poisson's equation are found to be

$$\begin{aligned} \psi(0,y) = & U(0,y) \\ & + \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \end{aligned} \quad (5.1.1)$$

$$+ \sum_{n=1}^{\infty} E_n \sin\left(\frac{n\pi y}{L}\right)$$

$$\frac{\partial\psi}{\partial x}(0,y) = -\sqrt{2K'} U(0,y)$$

$$+ \frac{1}{R t_{ox}} \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)]$$

$$- \sum_{n=1}^{\infty} \frac{n\pi}{L} E_n \tanh\left(\frac{n\pi d}{L}\right) \sin\left(\frac{n\pi y}{L}\right) \quad (5.1.2)$$

where the derivation of (5.1.2) makes use of (4.3.16) and (4.4.14). It is shown later in this chapter that if $\psi(0,y)$ is known then $\phi_n(y)$ can be calculated from it by applying the principle of current continuity. Since the E_n are the only unknowns in (5.1.1) and (5.1.2), if a choice is made for the E_n then $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ are immediately obtained from (5.1.1) and (5.1.2), and $\phi_n(y)$ is obtained from $\psi(0,y)$. It can then be checked to see if this choice of E_n provides for the one remaining condition of the continuity of normal displacement to see if the correct choice has been made. In practice, the infinite sums in (5.1.1) and (5.1.2) involving E_n are truncated to N terms, where N is 40 or 50 for short-channel devices, so choosing the E_n is really the equivalent of choosing a single N -dimensional vector $\underline{E}$, which could also be thought of as a single point in N -dimensional hyperspace. Thus the iterative process of sequentially applying the three conditions to the three functions amounts to a search through N -dimensional hyperspace to find the correct point that satisfies all three conditions. The difference between the iterative function technique and the Fourier analysis technique is that the iterative function technique leaves the three conditions in real space while the Fourier analysis technique transforms them into hyperspace, which is synonymous with Fourier space, and this provides for more elegant manipulations. Either way the progress of the iterative process is kept track of by assessing how good the present point in hyperspace is.

5.2 Applying Current Continuity to Calculate the Quasi-Fermi Potential

5.2a Calculating the Quasi-Fermi Potential From the Two-Dimensional Potential

To apply the above solution procedure, it is necessary to calculate $\phi_n(y)$ from $\psi(0,y)$ by applying the principle of current continuity. This is done using Gummel's algorithm [30] as developed for the MOSFET by De La Moneda [8] with some original modifications developed specifically for this work. The modifications are necessary because De La Moneda showed how $\phi_n(y)$ may be calculated from $\psi(x,y)$, but in the present case the sheet charge approximation is made so $\phi_n(y)$ must be calculated from $\psi(0,y)$.

The option pursued by De La Moneda of calculating $\phi_n(y)$ from $\psi(x,y)$ is not available if Poisson's equation is solved by analytical means. This is because to accurately calculate $\phi_n(y)$ from $\psi(x,y)$ the effect of mobile charge on $\psi(x,y)$ must be considered below the surface. In other words the inversion layer has a finite thickness and the sheet charge approximation cannot be made. But if the mobile charge is not restricted to an infinitely thin layer on the silicon surface then Poisson's equation assumes the highly nonlinear form of the Shockley-Poisson equation (2.1.1), which cannot be solved by the analytical techniques outlined in this thesis. Thus it is imperative for an analytical solution that the sheet charge approximation be used so that Poisson's equation is linearized

and hence soluble. The problem that then presents itself is that $\psi(0,y)$ will be reasonably reliable but $\psi(x,y)$ will deviate from its true value just below the surface by an amount which is almost negligible in terms of potential, but which will greatly perturb the carrier density due to its exponential dependence on $\psi(x,y)$. The solution to this problem is to deal with current continuity and the continuity of normal displacement in terms of the sheet charge approximation and hence they will only depend on $\psi(0,y)$. This is why it is necessary to find a method to calculate $\phi_n(y)$ from $\psi(0,y)$.

A method of this type has already been partly outlined by the derivation of the subthreshold current expression in Section 2.4. In principle, there is nothing that restricts this derivation to the subthreshold regime except that if the sheet charge approximation is used, and the device is above threshold, then the equations must be formulated in terms of $\psi(0,y)$ rather than $\psi(x,y)$ to avoid the inaccuracies in carrier densities that were just discussed. The derivation of the expression for $\phi_n(y)$ is begun by recalling that the development of the subthreshold drain current expression consisted of invoking one-dimensional current continuity on the drain current transport equation and making use of the integral expression for the mobile charge density Q_m . An expression for the quasi-Fermi potential function is obtained from this derivation by substituting (2.4.13) into (2.4.10) giving

$$z(y) = z(0) + (e^{-\beta V_{DS}} - 1) e^{-\beta(\phi_p + V_{SB})} \frac{\int_0^y \frac{dy}{\mu_n f}}{\int_0^L \frac{dy}{\mu_n f}} \quad (5.2.1)$$

Making use of the fact that

$$z(0) = e^{-\beta(\phi_p + V_{SB})} \quad (5.2.2)$$

$$\text{and } z(L) = e^{-\beta(\phi_p + V_{SB} + V_{DS})} \quad (5.2.3)$$

and rearranging the integrals yields

$$z(y) = z(0) \frac{\int_y^L \frac{dy}{\mu_n f}}{\int_0^L \frac{dy}{\mu_n f}} + z(L) \frac{\int_0^y \frac{dy}{\mu_n f}}{\int_0^L \frac{dy}{\mu_n f}} \quad (5.2.4)$$

The following definitions are now introduced

$$\lambda(y) = \frac{\int_0^y \frac{dy}{\mu_n f}}{\int_0^L \frac{dy}{\mu_n f}} \quad (5.2.5)$$

$$\lambda_C(y) = \frac{\int_y^L \frac{dy}{\mu_n f}}{\int_0^L \frac{dy}{\mu_n f}} \quad (5.2.6)$$

where it should be noted that $\lambda + \lambda_C = 1$, $0 \leq \lambda \leq 1$ and $0 < \lambda_C < 1$. Rearranging (5.2.4) and making use of (5.2.2), (5.2.3), (5.2.5) and (5.2.6), gives

$$\phi_n(y) = \phi_p + V_{SB} + \frac{V_{DS}}{2} - \frac{1}{\beta} \ln \cdot (e^{\frac{\beta}{2} V_{DS} + \ln \lambda_C} + e^{-\frac{\beta}{2} V_{DS} + \ln \lambda}) \quad (5.2.7)$$

Equation (5.2.7) is arranged in this way so that the logarithm term produces a variation about the average value of $\phi_n(y)$ which is $\phi_p + V_{SB} + V_{DS}/2$. This minimizes the problems caused by the function

$$g(x,y) = \ln (e^x + e^y) \quad (5.2.8)$$

that is used in (5.2.7), because when it is implemented on a digital computer it produces overflows and underflows if left in the form given by (5.2.8).

A practical method to compute $g(x,y)$ begins by checking to see if x is larger than or equal to y . If it is not then x and y are interchanged, which because of symmetry will not change the value of the function. Defining

$$z = x - y \quad (5.2.9)$$

where $z \geq 0$ because $x \geq y$, allows (5.2.8) to be rearranged to

$$g(x,y) = x + \ln (1 + e^{-z}) \quad (5.2.10)$$

Then expanding the logarithm in (5.2.10) into its Taylor series gives

$$g(x,y) = x + e^{-z} - \frac{e^{-2z}}{2} + \frac{e^{-3z}}{3} - \frac{e^{-4z}}{4} + \dots \quad (5.2.11)$$

Truncating the Taylor series to its first four terms and defining

$$u = e^{-z} \quad (5.2.12)$$

allows (5.2.11) to be rearranged to

$$q(x,y) = x + u \left(1 - \frac{u}{2} \left(1 - \frac{2}{3} u \left(1 - \frac{3}{4} u \right) \right) \right) \quad (5.2.13)$$

The function $q(x,y)$ is calculated to 12 significant digits by using (5.2.10) if $z < 8$ or using (5.2.13) if $z > 8$, if z is so large that it will cause u to underflow then u is simply set to zero. By following these rules $q(x,y)$ may be calculated without any numerical problems.

Substituting the definition for f from (2.4.4) into (5.2.5) gives

$$\lambda(y) = \frac{\int_0^y u_{n0} \int_0^\infty [e^{\beta\psi(x,y)} - 1] dx dy}{\int_0^L u_{n0} \int_0^\infty [e^{\beta\psi(x,y)} - 1] dx} \quad (5.2.14)$$

and the expression for $\lambda_C(y)$ is the same except that the limits on the first integral are y and L . Using (5.2.14), the related expression for $\lambda_C(y)$ and (5.2.7) it is possible to calculate $\phi_n(y)$ from $\psi(x,y)$. This is the version of Gummel's algorithm developed by De La Moneda [8].

5.2b Calculating the Quasi-Fermi Potential From the Surface Potential

To calculate $\phi_n(y)$ from $\psi(0,y)$ it is necessary to make the sheet charge approximation. This is done by rearranging (5.2.14) to

$$\lambda(y) = \frac{\int_0^y \frac{u_n e^{\beta\phi_n(y)} e^{-\beta\phi_n(y)} (-qn_i) \int_0^\infty [e^{\beta\psi(x,y)} - 1] dx}{\int_0^L \frac{u_n e^{\beta\phi_n(y)} e^{-\beta\phi_n(y)} (-qn_i) \int_0^\infty [e^{\beta\psi(x,y)} - 1] dx} dy}{\int_0^L \frac{u_n e^{\beta\phi_n(y)} e^{-\beta\phi_n(y)} (-qn_i) \int_0^\infty [e^{\beta\psi(x,y)} - 1] dx} dy} \quad (5.2.15)$$

where since

$$\frac{e^{\beta\phi_n(y)}}{e^{-\beta\phi_n(y)}} = 1 \quad (5.2.16)$$

this rearrangement does not change the values of the integrals. Rearranging (5.2.15) once more using the integral expression for Q_m (2.4.1) gives

$$\lambda(y) = \frac{\int_0^y \frac{u_n e^{\beta\phi_n} Q_m(\psi_s, \phi_n)}{\int_0^L \frac{u_n e^{\beta\phi_n} Q_m(\psi_s, \phi_n)}{dy}} dy}{\int_0^L \frac{u_n e^{\beta\phi_n} Q_m(\psi_s, \phi_n)}{dy}} \quad (5.2.17)$$

where $\psi_s = \psi(0, y)$.

Substitution of (4.2.4) and (4.2.5) into (4.2.3) allows the sheet charge expression for the mobile charge density to be written as

$$Q_m(\psi_s, \phi_n) = -\sqrt{\frac{2q\epsilon_s N_B}{\beta}} ([e^{\beta(\psi_s - \phi_n - \phi_p)} + \beta\psi_s - 1]^{1/2} - (\beta\psi_s - 1)^{1/2}) \quad (5.2.18)$$

The difficulty with using (5.2.18) in (5.2.17) is that when Q_m becomes very small the difference taken in (5.2.18) causes Q_m to become very inaccurate because of the necessity of using a finite word length in the computations. This poses a major problem for calculating λ because the integrals in (5.2.17) depend on Q_m^{-1} , hence high accuracy is required when Q_m is small. To overcome this difficulty (5.2.18) is rearranged to

$$Q_m(\psi_s, \phi_n) = -\sqrt{\frac{2 q \epsilon_s n_i}{\beta}} e^{\frac{\beta}{2}(\psi_s - \phi_n)} f(T) \quad (5.2.19)$$

$$\text{where } f(T) = \sqrt{1+T} - \sqrt{T} \quad (5.2.20)$$

$$T = (\beta \psi_s - 1) e^{-\beta(\psi_s - \phi_n - \phi_p)} \quad (5.2.21)$$

To achieve maximum accuracy the function $f(T)$ is calculated using the following rules

$$\text{If } T < 10^{-14} \text{ then } f(T) = 1 \quad (5.2.22)$$

$$\text{If } 10^{-14} \leq T \leq 10^8 \text{ then } f(T) = \sqrt{1+T} - \sqrt{T} \quad (5.2.23)$$

$$\text{If } T > 10^8 \text{ then } f(T) = \frac{1}{2\sqrt{T}} \quad (5.2.24)$$

where (5.2.24) is obtained from (5.2.20) using the binomial approximation.

Substituting (5.2.19) into (5.2.17) gives

$$\lambda(y) = \frac{\int_0^y \frac{e^{-\frac{\beta}{2}(\psi_s + \phi_n)}}{\mu_n f(T)} dy}{\int_0^L \frac{e^{-\frac{\beta}{2}(\psi_s + \phi_n)}}{\mu_n f(T)} dy} \quad (5.2.25)$$

The exponential in the integrand will cause underflows if it is left in this form so the numerator and denominator of (5.2.25) are multiplied by a constant to yield

$$\lambda(Y) = \frac{\int_0^Y e^{-\frac{\beta}{2}(\psi_s - \psi_{ss} + \phi_n - \phi_{nD})} \frac{dy}{u_n f(T)}}{\int_0^L e^{-\frac{\beta}{2}(\psi_s - \psi_{ss} + \phi_n - \phi_{nD})} \frac{dy}{u_n f(T)}} \quad (5.2.26)$$

$$\text{where } \psi_{ss} = \psi(0,0) \quad (5.2.27)$$

$$\phi_{nD} = \phi_n(L) \quad (5.2.28)$$

In a similar manner $\lambda_C(Y)$ is found to be

$$\lambda_C(Y) = \frac{\int_Y^L e^{-\frac{\beta}{2}(\psi_s - \psi_{ss} + \phi_n - \phi_{nD})} \frac{dy}{u_n f(T)}}{\int_0^L e^{-\frac{\beta}{2}(\psi_s - \psi_{ss} + \phi_n - \phi_{nD})} \frac{dy}{u_n f(T)}} \quad (5.2.29)$$

The drain current expression developed in Section 2.4 for the subthreshold model can be modified in the same way as $\phi_n(y)$ so that I_D can be calculated from $\psi(0,y)$ and $\phi_n(y)$. This provides a valid drain current expression for the general model because (2.4.14) makes no assumptions that restricts its use to subthreshold. Modifying the integral in the denominator of (2.4.14) as was done in (5.2.17), applying the sheet charge expression for Q_m (5.2.19), and multiplying by constants to prevent underflows yields the drain current expression

$$I_D = \frac{W}{\beta} \sqrt{\frac{2q\epsilon_s n_i}{\beta}} e^{\frac{-\beta(\phi_p + V_{SB} - \frac{\psi_{ss} + \phi_{nD}}{2})}{1 - e^{-\beta V_{DS}}}} \int_0^L \frac{e^{\frac{-\beta(\psi_s - \psi_{ss} + \phi_n - \phi_{nD})}{2}}}{u_n f(T)} dy \quad (5.2.30)$$

The mobility u_n indicated in (5.2.26), (5.2.29) and (5.2.30) can be either a constant or a field dependent mobility. If it is field dependent then the Coen and Muller [58] mobility model is used which was determined from measurements on a resistive gate MOSFET for carriers in the inversion layer. It is of the form

$$u_n = \frac{u_{no}}{\left(1 + \frac{E_x}{E_{cx}}\right) \left[1 + \left(\frac{E_y}{E_{cy}}\right)^\alpha\right]^{1/\alpha}} \quad (5.2.31)$$

where for electrons in p-type (100) silicon

$$u_{no} = 850 \text{ cm}^2/\text{Vs} \quad (5.2.32)$$

$$E_{cx} = 7 \times 10^5 \text{ V/cm}$$

$$E_{cy} = 8.7 \times 10^3 \text{ V/cm}$$

$$\alpha = 2.9$$

The magnitude of the electric field components E_x and E_y are calculated from

$$E_x = \frac{V_{GB} - \psi(0, y)}{R t_{ox}} \quad (5.2.33)$$

$$E_y = \left| \frac{\partial \psi}{\partial y}(0, y) \right| \quad (5.2.34)$$

where E_x is set to zero if $V_{GB} < \psi(0,y)$ because the Coen and Muller mobility model is not strictly valid for field reversal but it is used there anyway by restricting the field dependence to E_y .

The method by which the quasi-Fermi potential $\phi_n(y)$ is calculated from the surface potential $\psi(0,y)$ has now been completely described. It is done using (5.2.7), (5.2.26) and (5.2.29) in an iterative fashion because they have a transcendental dependence on $\phi_n(y)$. The initial guess that is used for $\phi_n(y)$ is its prevailing estimate at that time and usually only two or three iterations are required for $\phi_n(y)$ to converge. The convergence is monitored by calculating I_D using (5.2.30) and when I_D converges $\phi_n(y)$ converges. Numerical difficulties with overflows and underflows are avoided by calculating the function $\ln(e^x + e^y)$ using (5.2.10) and (5.2.13). The integrals in (5.2.26) and (5.2.29) are done numerically using a 10 point Gaussian quadrature formula [59,60] which is both accurate and efficient. Thus the principle of current continuity has been cast in a form where it can be made use of in the iterative function technique. This method is hereafter referred to as Gummel's algorithm after the originator of the concept [30].

5.3 Employing the Continuity of Normal Displacement to Calculate the Surface Potential

It has been described in this chapter how Poisson's equation and the principle of current continuity are used as part of the iterative function technique. Poisson's equation permits the calculation of $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ from the prevailing estimate of the Fourier coefficients E_n , and current continuity facilitates the calculation of $\phi_n(y)$ from $\psi(0,y)$. What remains to be described is how the continuity of normal displacement across the silicon-silicon dioxide interface is enforced and how this fits into the iterative function technique.

If the sheet charge approximation is made then the continuity of normal displacement places the following restriction on the three functions $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$.

$$C_{ox} [V_{GB} - \psi(0,y)] + Q_m(\psi(0,y), \phi_n(y)) + \epsilon_s \frac{\partial\psi}{\partial x}(0,y) = 0 \quad (5.3.1)$$

This is the same condition required by (4.2.2) in the calculation of the surface potential used by the single parabola boundary condition. Equation (5.3.1) can be used to determine the value of any one of the three functions at a point $(0,y)$ provided that the values of the other two functions are known. It was chosen to calculate $\psi(0,y)$ from the prevailing estimates of $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$ because this is directly related to the calculation of the boundary potentials U_s and ψ_s dealt with in Chapter 4 about

which a great deal is known. It is also the only choice that makes sense in the iterative sequence of calculations that is described later in this section.

Substituting the sheet charge expression for Q_m (4.2.3) into (5.3.1) gives

$$C_{ox} (V_{GB} - \psi_s) - C_K \{ [e^{\beta(\psi_s - \phi)} + \beta\psi_s - 1]^{1/2} - (\beta\psi_s - 1)^{1/2} \} + \epsilon_s \psi_x = 0 \quad (5.3.2)$$

$$\text{where } \psi_x = \frac{\partial \psi}{\partial x} (0, y) \quad (5.3.3)$$

Differentiating (5.3.2) with respect to ψ_s permits the construction of the following Newton-Raphson iterative equation

$$\psi'_s = \psi_s + \frac{C_{ox} (V_{GB} - \psi_s) - C_K \{ [e^{\beta(\psi_s - \phi)} + \beta\psi_s - 1]^{1/2} - (\beta\psi_s - 1)^{1/2} \} + \epsilon_s \psi_x}{C_{ox} + \frac{\beta}{2} C_K \left\{ \frac{e^{\beta(\psi_s - \phi)} + 1}{[e^{\beta(\psi_s - \phi)} + \beta\psi_s - 1]^{1/2}} - \frac{1}{(\beta\psi_s - 1)^{1/2}} \right\}} \quad (5.3.4)$$

where ψ'_s is the next value for ψ_s in the iterative sequence. Note the similarity of (5.3.4) and (4.2.10) which is due to the fact that both are Newton-Raphson equations used to calculate ψ_s from a sheet charge expression of the continuity of normal displacement. The only difference is in the expression for ψ_x used in (4.2.10). It is possible to derive equations for the above and below threshold initial guesses to be used with (5.3.4) just as was done for (4.2.10). But since (5.3.4) is used with

the iterative function technique this is found to be unnecessary because a prevailing estimate for $\psi(0,y)$ is always available that is good enough to be used as an initial guess. Thus the continuity of normal displacement is enforced by using the prevailing value of $\psi(0,y)$ as an initial guess, and iterating using (5.3.4) usually no more than twice so that an improved estimate of $\psi(0,y)$ is obtained.

5.4 The Algorithms

Now that Poisson's equation, current continuity, and the continuity of normal displacement have been cast in forms suitable for use by the iterative function technique, the algorithms utilized by this technique may be described. Two workable algorithms were identified and the flowcharts for these algorithms are shown in Figure 5.4.1.

Both algorithms begin with initial guesses for $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$. They then update $\phi_n(y)$ using Gummel's algorithm to calculate it from $\psi(0,y)$, and update $\psi(0,y)$ using the continuity of normal displacement. At this point in the algorithms the mobile charge density $Q_m(y)$ is the correct order of magnitude, although not exactly the correct value, because the continuity of normal displacement has just been applied. Thus this is a good point at which to calculate $Q_m(y)$ from $\psi(0,y)$ and $\phi_n(y)$, then calculate $V_{SC}(y)$ from

$$V_{SC}(y) = \frac{Q_{mLC}(y) - Q_m(y)}{C_{ox}} \quad (5.4.1)$$

Initial
Guesses

(a)

Initial
Guesses

(b)

Figure 5.4.1. (a) Algorithm 1 and (b) Algorithm 2 for the iterative function technique.

where (5.4.1) is obtained by substituting (4.4.2) into (4.4.4). Upon Fourier analyzing $V_{SC}(y)$ using (4.4.38) the Fourier coefficients E_n , which can also be denoted by the Fourier vector E , are determined. At this point the two algorithms become different. Algorithm 1 uses E to calculate only $\partial\psi/\partial x(0,y)$ while Algorithm 2 uses E to calculate both $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$. The difference results from the fact that since $\psi(0,y)$ is updated by invoking the continuity of normal displacement it is not essential to update it using the latest estimate of E . It is sufficient to only update $\partial\psi/\partial x(0,y)$ as is done by Algorithm 1, but it is still permissible to update both $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ using E as is done by Algorithm 2. This difference in the two algorithms has essentially no effect on their ability to converge to a solution as is discussed later in this chapter after certain details of the iterative function technique are presented.

The first such detail is how the initial guesses for $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$ are obtained. These are generated by starting with a guess for $\psi(0,y)$ of

$$\psi(0,y) = \frac{[\psi(0,0) - U(0,0)] \sinh(\frac{L-y}{d}) + [\psi(0,L) - U(0,L)] \sinh(y/d)}{\sinh(L/d)} + U(0,y) \quad (5.4.2)$$

This function was chosen because it is smooth, has the correct boundary potentials $\psi(0,0)$ and $\psi(0,L)$, is always greater than the long-channel surface potential $U(0,y)$, and its short-channel component falls off exponentially from each end of the channel. Its functional form was inspired by the y dependence of $V(x,y)$ (see (2.2.21)), because the hyperbolic sine terms in this expression are arranged so that one of them vanishes at each boundary (i.e. $y = 0$ and $y = L$), making it easy to choose the value of the function at these points. Using this initial guess for $\psi(0,y)$ implies an E vector from which an initial guess for $\partial\psi/\partial x(0,y)$ may be obtained. The Fourier coefficients E_n are extracted from (5.4.2) by setting x to zero in (4.4.34) giving

$$Z(0,y) = \sum_{n=1}^{\infty} E_n \sin\left(\frac{n\pi y}{L}\right) \quad (5.4.3)$$

Employing the orthogonality relationship (4.4.37) and noting that

$$Z(0,y) = \psi(0,y) - U(0,y) - V(0,y) \quad (5.4.4)$$

(5.4.3) is rearranged to

$$E_n = \frac{2}{L} \int_0^L [\psi(0,y) - U(0,y) - V(0,y)] \sin\left(\frac{n\pi y}{L}\right) dy \quad (5.4.5)$$

Equation (5.4.5) is used to extract the E_n from the initial guess for $\psi(0,y)$ by means of Fourier analysis. Once the E_n are calculated they are used in (5.1.2) to calculate an initial guess for $\partial\psi/\partial x(0,y)$.

Since approximations now exist for $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ the continuity of normal displacement can be applied to calculate $\phi_n(y)$. This involves rearranging (5.3.2) to

$$\phi_n = \psi_s - \frac{1}{\beta} \ln \left[\frac{N_B}{n_i} \left(\frac{C_{ox}}{C_K} (V_{GB} - \psi_s)^{1+\epsilon_s \psi_x} + \sqrt{\beta \psi_s - 1} \right)^2 - \beta \psi_s + 1 \right] \quad (5.4.6)$$

which is a closed form expression for $\phi_n(y)$ as a function of $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ (i.e. ψ_s and ψ_x). In the event that the argument of the logarithm becomes less than 1 then evaluating the logarithm is dispensed with and ϕ_n is set equal to ψ_s .

5.5 Patterson's Method of Evaluating Fourier Integrals

5.5a Discretizing the Fourier Integrals

Although the concept of Fourier analysis has been discussed in this chapter and has been used explicitly in (4.4.38) and (5.4.5), nothing has been said about evaluating Fourier integrals of the form

$$I_{cn} = \frac{2}{L} \int_0^L f(t) \cos \left(\frac{n\pi t}{L} \right) dt, \quad n = 0, 1, 2, \dots \quad (5.5.1)$$

$$I_{sn} = \frac{2}{L} \int_0^L f(t) \sin \left(\frac{n\pi t}{L} \right) dt, \quad n = 1, 2, 3, \dots \quad (5.5.2)$$

which are the integrals used to calculate the coefficients for a half-range Fourier sine or cosine series. Note that the cosine

integral given by (5.5.1) has not yet been encountered but is needed in Chapter 6 so it is discussed here. Integrals of this type are used in (4.4.38) and (5.4.5) but cannot be evaluated analytically so they must be evaluated numerically. This is a very difficult numerical problem because n does not have to be very large, say 5 or 10, for traditional numerical integration methods like Simpson's Rule or Gaussian Quadrature to fail utterly. The reason for this failure is that for a large n the integrands in (5.5.1) and (5.5.2) are highly oscillatory. Fortunately, Fourier analysis is such an important concept to so many fields of endeavour that a great deal of research has been done on the numerical evaluation of Fourier integrals. The most efficient of these methods was chosen from the literature and made even better to produce a truly powerful technique for evaluating (5.5.1) and (5.5.2).

To begin the explanation of this technique (5.5.1) and (5.5.2) must be transformed from the interval $[0, L]$ to $[-1, 1]$ and be specialized to either n odd or n even. Thus the cosine integral (5.5.1) becomes:

$$I_{C2k} = (-1)^k \int_{-1}^1 g(u) \cos(k\pi u) du, \quad k=0, 1, 2, \dots \quad (5.5.3)$$

$$I_{C2k-1} = (-1)^k \int_{-1}^1 g(u) \sin\left(\frac{2k-1}{2} \pi u\right) du, \quad k=1, 2, 3, \dots \quad (5.5.4)$$

where (5.5.3) is the case for n even (i.e. $n = 2k$) and (5.5.4) is the case for n odd (i.e. $n = 2k-1$). Similarly for the sine integral (5.5.2):

$$I_{s2k} = (-1)^k \int_{-1}^1 g(u) \sin(k\pi u) du, \quad k=1,2,3 \dots \quad (5.5.5)$$

$$I_{s2k-1} = (-1)^{k+1} \int_{-1}^1 g(u) \cos\left(\frac{2k-1}{2} \pi u\right) du, \quad k=1,2,3 \dots \quad (5.5.6)$$

$$\text{where } g(u) = f\left(\frac{1}{2}(u+1)\right) \quad (5.5.7)$$

The most powerful numerical technique available in the literature to evaluate Fourier integrals normalized to the interval $[-1,1]$ was originated by Patterson [61]. His method depends upon expanding $g(u)$ using Chebyshev polynomials, then expanding again using Legendre polynomials, and finally integrating the product of the Legendre polynomials and the sines or cosines which yields spherical Bessel functions. This produces the two relatively simple expressions

$$\int_{-1}^1 g(u) \cos(wu) du = \sum_{i=0}^{[N_c/2]} a_{2i} \sum_{k=0}^i (-1)^k (4k+1) R_{2k,2i} j_{2k}(w) \quad (5.5.8)$$

$$\int_{-1}^1 g(u) \sin(wu) du = \sum_{i=0}^{[N_c-1/2]} a_{2i+1} \sum_{k=0}^i (-1)^k (4k+3) R_{2k+1,2i+1} j_{2k+1}(w) \quad (5.5.9)$$

where the a_i are the coefficients of the Chebyshev expansion, N_C+1 is the number of terms in the Chebyshev expansion, the $j_k(w)$ are spherical Bessel functions and the R_{ki} are given by

$$R_{ki} = \int_{-1}^1 P_k(u) T_i(u) du \quad (5.5.10)$$

where the $P_k(u)$ are the Legendre polynomials and the $T_i(u)$ are the Chebyshev polynomials. Note that the square brackets around $[N_C/2]$ and $[(N_C-1)/2]$ imply that the result from the division is to be truncated.

5.5b The Chebyshev Coefficients

The Chebyshev coefficients a_i of the Chebyshev expansion [62]:

$$g(u) = \sum_{i=0}^{N_C} a_i T_i(u) \quad (5.5.11)$$

are given by the collocation expressions

$$a_i = \frac{2}{N_C} \sum_{j=0}^{N_C} g(u_j) T_i(u_j) \text{ for } 1 \leq i \leq N_C - 1 \quad (5.5.12)$$

$$a_i = \frac{1}{N_C} \sum_{j=0}^{N_C} g(u_j) T_i(u_j) \text{ for } i = 0 \text{ or } N_C \quad (5.5.13)$$

where the collocation points u_j are defined by

$$u_j = -\cos\left(\frac{\pi j}{N_C}\right), \quad j = 0, 1, 2 \dots N_C \quad (5.5.14)$$

and N_C+1 is the number of collocation points. The double primes on the summation signs in (5.5.12) and (5.5.13) indicate that the first and last terms are to be divided by 2. The values of the Chebyshev polynomials $T_i(u)$ are easily generated using the recurrence relation

$$T_{i+1}(u) = 2u T_i(u) - T_{i-1}(u) \quad (5.5.15)$$

$$\text{where } T_0(u) = 1 \quad (5.5.16)$$

$$T_1(u) = u \quad (5.5.17)$$

The reason the function $g(u)$ is first expanded into Chebyshev polynomials instead of proceeding directly to a Legendre polynomial expansion is that the collocation expressions (5.5.12) and (5.5.13) calculate the Chebyshev coefficients a_i so efficiently. This is because to generate N_C+1 coefficients requires only N_C+1 function evaluations of $g(u)$.

5.5c The R_{ki}

The R_{ki} are easily obtained from (5.5.10) by first calculating R_{0i} and R_{1i} . Since

$$P_0(u) = 1 \quad (5.5.18)$$

$$R_{0i} = \int_{-1}^1 T_i(u) du \quad (5.5.19)$$

which transforms to

$$R_{0i} = \int_0^\pi \cos(i\theta) \sin \theta d\theta \quad (5.5.20)$$

yielding

$$R_{0i} = \frac{2}{1 - i^2} \text{ for } i \text{ even} \quad (5.5.21)$$

$$0 \quad \text{for } i \text{ odd}$$

Similarly

$$R_{1i} = \frac{2}{4 - i^2} \text{ for } i \text{ odd} \quad (5.5.22)$$

$$0 \quad \text{for } i \text{ even}$$

The first two rows of the R_{ki} matrix are generated by (5.5.21) and (5.5.22). All subsequent rows are obtained from the first two rows using the following recurrence relation developed by Patterson [61].--

$$R_{k+2, i} = \left[\frac{i^2 - k^2}{i^2 - (k+3)^2} \right] R_{ki} \quad (5.5.23)$$

Thus the R_{ki} can be calculated with very little effort.

5.5d Calculating the Spherical Bessel Functions

The calculation of the spherical Bessel functions $j_k(w)$ must be done with some care because the forward recurrence relation

$$j_{k+1}(w) = \frac{2k+1}{w} j_k(w) - j_{k-1}(w) \quad (5.5.24)$$

becomes unstable for $k > w + 1$. But it can be used for $k \leq w + 1$ by starting with the expressions for $j_0(w)$ and $j_1(w)$

$$j_0(w) = \frac{\sin w}{w} \quad (5.5.25)$$

$$j_1(w) = \frac{\sin w}{w^2} - \frac{\cos w}{w} = \frac{j_0(w) - \cos w}{w} \quad (5.5.26)$$

Note that by using the second form of (5.5.26) it is possible to calculate $j_0(w)$ and $j_1(w)$ with a single sine and a single cosine evaluation.

For $k > w + 1$ it is necessary to use a backwards recurrence relation as outlined by Littlewood and Zakian [63] whose discussion is based on the continued fraction work of Blanch [64]. This approach consists of rearranging the forward recurrence relation (5.5.24) to

$$\frac{j_{k+1}}{j_k} = \frac{2k+1}{w} - \frac{j_{k-1}}{j_k} \quad (5.5.27)$$

and defining

$$R_k = \frac{j_{k-1}}{j_k} \quad (5.5.28)$$

giving

$$R_k = \frac{2k+1}{w} - \frac{1}{R_{k+1}} \quad (5.5.29)$$

By substituting (5.5.29) back into itself repeatedly a continued fraction of the form

$$R_k = \frac{2k+1}{\omega} - \frac{1}{\frac{2(k+1)+1}{\omega} - \frac{1}{\frac{2(k+2)+1}{\omega} - \frac{1}{\frac{2(k+3)+1}{\omega} - \text{etc.}}}} \quad (5.5.30)$$

is created. This continued fraction is evaluated using the technique developed by Blanch [64] which involves the iterative calculation of R_m using

$$A_m = b_m A_{m-1} - A_{m-2} \quad (5.5.31)$$

$$B_m = b_m B_{m-1} - B_{m-2} \quad (5.5.32)$$

$$R_m = A_m/B_m \quad (5.5.33)$$

$$\text{where } b_m = \frac{2(k+m)+1}{\omega} \quad (5.5.34)$$

$$A_{-1} = 1 \quad (5.5.35)$$

$$B_{-1} = 0 \quad (5.5.36)$$

$$A_0 = b_0 \quad (5.5.37)$$

$$B_0 = 1 \quad (5.5.38)$$

R_k is then obtained in the limit

$$R_k = \lim_{m \rightarrow \infty} R_m \quad (5.5.39)$$

Typically R_k is calculated to 14 significant digits (i.e. double precision on an IBM 3081KX computer) in less than 10 iterative steps. Once R_k is found all the R_k back to the k at which the forward recurrence relation began to fail are calculated using (5.5.29) as a backwards recurrence relation. The ratios R_k are then used to calculate the $j_k(w)$ using the last $j_k(w)$ found with the forward recurrence relation. Thus by employing the forward and backwards recurrence relations any number of $j_k(w)$ may be calculated with a single sine and a single cosine evaluation and a relatively small number of multiplications and divisions.

5.5e A More Efficient Form of Patterson's Method

A more efficient form of Patterson's method was developed for use with this thesis work. The increased efficiency is achieved by noting that if w is known ahead of time, as it is for the present application, then the second sums in (5.5.8) and (5.5.9) may be calculated and stored as a set of coefficients because they are independent of the function $g(u)$. Thus by defining the coefficient matrices

$$S_{Cin} = (-1)^n (-1)^{(n/2)} \sum_{k=0}^i (-1)^k (4k+1) R_{2k,2i} j_{2k} \left(\frac{n\pi}{2} \right) \quad (5.5.40)$$

$$S_{Ci,0} = \frac{2}{1-(2i)^2} \quad (5.5.41)$$

$$S_{Sin} = (-1)^n (-1)^{(n/2)} \sum_{k=0}^i (-1)^k (4k+3) R_{2k+1,2i+1} j_{2k+1} \left(\frac{n\pi}{2} \right) \quad (5.5.42)$$

where the brackets around $(n/2)$ imply a truncation, and applying (5.5.8) and (5.5.9), allows (5.5.3) to (5.5.6) to be written as

$$I_{c2k} \approx \sum_{i=0}^{N_C/2} a_{2i} S_{ci,2k} \quad (5.5.43)$$

$$I_{c2k-1} \approx \sum_{i=0}^{\left[\frac{N_C-1}{2} \right]} a_{2i+1} S_{si,2k-1} \quad (5.5.44)$$

$$I_{s2k} \approx \sum_{i=0}^{\left[\frac{N_C-1}{2} \right]} a_{2i+1} S_{si,2k} \quad (5.5.45)$$

$$I_{s2k-1} \approx - \sum_{i=0}^{\left[N_C/2 \right]} a_{2i} S_{ci,2k-1} \quad (5.5.46)$$

This simplification results in what is essentially a quadrature method for evaluating Fourier integrals. The difference between the new method and the usual quadrature methods is that the function values are replaced by the Chebyshev coefficients of the function. The new method is very fast and efficient. It requires N_C+1 function evaluations and $(N_C+1)^2$ multiplications to calculate the Chebyshev coefficients a_i , and thereafter of the order of $N_C/2$ multiplications to evaluate each Fourier integral. The price paid for this speed is that a lot of storage is required. On the IBM 3081KX computer used for these calculations N_C was limited to approximately 40 because larger values of N_C caused the spherical Bessel functions to underflow the computer.

The number of Fourier coefficients N that were needed was at most 50. With an N_C of 40 and an N of 50 there were 2079 elements of the coefficient matrices S_{Cin} and S_{Sin} to be stored. This required 2079 FORTRAN DATA statements to store the coefficients in double precision which were written automatically by the same program that generates the coefficients so as to make the programming of the coefficients practical. The large amount of storage made compilation times long, but the speed offered by the high speed Fourier analysis technique described by equations (5.5.40) to (5.5.46) was deemed to be worth it, because the ultimate objective of this thesis is a high speed simulator for the short-channel MOSFET.

5.6 The Performance of the Iterative Function Technique

Since the numerical evaluation of the integrals in (5.4.5) and (4.4.38) requires that $V_{SC}(y)$, $\psi(0,y)$, $U(0,y)$ and $V(0,y)$ be expanded in terms of Chebyshev polynomials, and the iterative function technique requires the storage of the functions $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$ on a computer, it was decided to store all these functions of y as Chebyshev expansions. This makes good sense because Chebyshev expansions are widely used for function storage. Another advantage is that $\partial\psi/\partial y(0,y)$ is easily obtained from the Chebyshev expansion of $\psi(0,y)$ so that it can be used to calculate the field dependent mobility. This is done by taking the Chebyshev expansion of $\psi(0,y)$

$$\psi(0, y) = \sum_{i=0}^{N_c} a_i T_i(u) \quad (5.6.1)$$

$$\text{where } u = \frac{2y}{L} - 1 \quad (5.6.2)$$

and differentiating it to obtain

$$\frac{\partial \psi}{\partial y}(0, y) = \frac{2}{L} \sum_{i=1}^{N_c} a_i \frac{dT_i}{du}(u) \quad (5.6.3)$$

The dT_i/du are calculated by differentiating the recurrence relation for the T_i (5.5.15) giving

$$\frac{dT_{i+1}}{du} = 2 \left(T_i + u \frac{dT_i}{du} \right) - \frac{dT_{i-1}}{du} \quad (5.6.4)$$

$$\text{where } \frac{dT_0}{du} = 0 \quad (5.6.5)$$

$$\frac{dT_1}{du} = 1 \quad (5.6.6)$$

Having recorded the functions $\psi(0, y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0, y)$ as Chebyshev expansions, the iterative function technique becomes a procedure to calculate the correct Chebyshev coefficients for these expansions. Thus as the calculations progress in algorithms 1 and 2 the vectors of Chebyshev coefficients as well as the vector of Fourier coefficients E_n are constantly being updated. It has been found that a useful technique for keeping track of the progress of the iterative function technique is to

plot the drain current I_D , calculated from (5.2.30), versus iteration number. Plots of this type are shown in Figures 5.6.1 and 5.6.2 for algorithms 1 and 2 respectively. The (a) part of each figure is for a constant mobility and the (b) part is for a field dependent mobility. The bias conditions are $V_{GS} = 5$ V, $V_{DS} = 1$ V and $V_{SB} = 0$ and the device used is Device 1 whose parameters are given in Table 4.3.1.

The predominant feature of the I_D versus iteration number plots is that the iteration proceeds as a damped oscillation. This is due to the highly nonlinear nature of the problem that is being solved. Because of the oscillatory nature of the iteration it is difficult to define convergence, but it appears that for this device and bias conditions 130 iterations are required to obtain convergence if the mobility is constant and 180 iterations are required if the mobility is field dependent. The number of iterations needed to reach convergence is apparently independent of the algorithm employed. Although not shown by these figures, it has been found that the number of iterations is less for a smaller drain voltage, dropping to around 10 iterations for a V_{DS} of 0.1 V.

5.7 Why the Iterative Function Technique Is Too Slow

The execution times required to obtain convergence for the cases depicted in Figures 5.6.1 and 5.6.2 are 46 CPU seconds for a constant mobility and 75 CPU seconds for a field dependent mobility on an IBM 3081KX computer. For the same device and bias conditions the lower level of MINIMOS, that uses a

(a)

(b)

Figure 5.6.1. Drain current versus iteration number for Device 1 using Algorithm 1 where the mobility is constant in (a) and field dependent in (b).

(a)

(b)

Figure 5.6.2. Drain current versus iteration number for Device 1 using Algorithm 2 where the mobility is constant in (a) and field dependent in (b).

one-dimensional current continuity equation and hence is directly comparable to the analytical model, requires 3.6 CPU seconds. In general MINIMOS requires between 2 and 60 CPU seconds depending on the device structure and bias conditions. Clearly when the analytical solution is completed using the iterative function technique the resulting model is in general slower than the numerical solution of the MINIMOS model. Thus the iterative function technique has failed to meet the objective defined in Chapter 1 of developing an analytical model of the short-channel MOSFET which is faster than comparable numerical models. There are three reasons for this:

The first is that the iterative function technique does not take full advantage of the analytical formulation of the equations to be solved. It goes to numerical methods too soon where it is more appropriate to combine the equations into a single equation by analytical techniques before numerical methods are applied.

The second reason is that there is no mechanism to accelerate the convergence of the iterative function technique. The most obvious method to do this is to deal with the problem as a system of N nonlinear equations in N unknowns and apply the Newton and quasi-Newton methods described in the next section. Attempts to do this failed, apparently because more analytical manipulation is required to make the problem amenable to this technique.

The third reason is that numerical noise is generated by calculating $\psi(0,y)$ and $\partial\psi/\partial x(0,y)$ from the Fourier series given by (5.1.1) and (5.1.2), and by recording $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$ as Chebyshev expansions. The noise is due to the ripple in the representation of these functions caused by truncating the Fourier and Chebyshev series. This level of noise is only a problem because the equations are so highly nonlinear. The major nonlinearity is the exponential dependence of $Q_m(y)$ on $\psi(0,y)$ and $\phi_n(y)$ (see (4.2.3)), which amplifies the noise in $\psi(0,y)$ and $\phi_n(y)$. The noise in $Q_m(y)$ is transferred to $V_{SC}(y)$ and E , and is then fed back into the iterative system. This noise hampers convergence because it produces an extra randomness in the iterative search.

The shapes of the two-dimensional potential functions $\psi(0,y)$, $U(x,y)$, $V(x,y)$ and $Z(x,y)$ determined using algorithm 1 are shown in Figures 5.7.1 for Device 1 biased at $V_{GS} = 5$ V, $V_{DS} = 1$ V and $V_{SB} = 0$. They are plotted over the rectangular region under consideration (see Figure 4.1.1) and the perspective is looking from the bulk up to the surface. The effect of noise on the two-dimensional potential is best demonstrated by changing the perspective of Figure 5.7.1d and viewing $Z(x,y)$ from the surface as is shown in Figure 5.7.2a. Because the drain voltage is small and a relatively large number of terms (i.e. 41) has been used in the Chebyshev expansions, the effect of noise is only evident in the ripple in the lower left hand part of Figure 5.7.2a. But if the number of terms in the Chebyshev

Figure 5.7.1. Three-dimensional plots of (a) $\psi(x,y)$, (b) $U(x,y)$, (c) $V(x,y)$ and (d) $Z(x,y)$ as viewed from the substrate, where with the mobility constant these potentials were solved for using Algorithm 1.

Figure 5.7.2. Three-dimensional plots of $Z(x,y)$ as viewed from the silicon surface where in (a) 41 Chebyshev terms were used and in (b) 21 Chebyshev terms were used.

expansions is reduced to 21 then the effect of noise is quite evident as displayed in Figure 5.7.2b. Once this noise is amplified by the exponential dependence of $Q_m(y)$ on $\psi(0,y)$ its effect is quite pronounced.

5.8 Summary

The iterative function technique is not the best method to use to solve for the Fourier coefficients E_n , but a great deal has been learned from it. For instance, it is now known what the plots of $\psi(x,y)$, $U(x,y)$, $V(x,y)$ and $Z(x,y)$ look like (see Figures 5.7.1 and 5.7.2). From these plots it is seen that the rough shapes assumed by $V(x,y)$ and $Z(x,y)$ work together when added to $U(x,y)$ to produce the smooth shape of $\psi(x,y)$. The only anomaly in $\psi(x,y)$ is the little bump near the point $(d,0)$ that is the result of neglecting the short-channel displacement crossing the bottom boundary of the box in boundary condition (4.1.5). The shape of $Z(x,y)$ is also quite informative as the deviation of $Z(0,y)$ from zero (see Figure 5.7.2) is directly proportional to the short-channel component of mobile charge $Q_{MSC}(y)$ since $V_{SC}(y)$ is the driving function of Laplace problem 2 (see equations (4.4.16) to (4.4.20)). Thus viewing $Z(x,y)$ from the surface is actually a map of the presence of short-channel charge and hence short-channel effects. Not surprisingly, the short-channel charge is concentrated near the source and drain.

In summary, what has been learned from the iterative function technique is that calculating the E_n is the equivalent

of searching for a particular point in hyperspace by applying the conditions dictated by Poisson's equation, current continuity, and the continuity of normal displacement. The slowness of the iterative function technique is unsatisfactory and to do better than this the three conditions must be combined in a more intelligent fashion, and better numerical techniques must be used. A Fourier analysis technique that meets these specifications is presented in the next chapter.

CHAPTER 6

THE FOURIER ANALYSIS TECHNIQUE

6.1 Transformation of the Problem into the Fourier Domain

In the last chapter it was found that the iterative function technique was capable of calculating the Fourier coefficients E_n , but that when it was implemented on a computer it was too slow. A faster solution method can be achieved by combining the conditions imposed by Poisson's equation, current continuity and the continuity of normal displacement into a single equation. The transformation of this equation into Fourier space facilitates an elegant and rapid method of calculating the Fourier coefficients E_n . This technique is the usual way to complete the solution of a Laplace problem. An example is the Fourier analysis of $V(x,0)$ and $V(x,L)$ in (2.2.22) and (2.2.23) in Laplace problem 1. The difference between these equations and the corresponding equation in Laplace problem 2 (4.4.38), is that (4.4.38) has a transcendental dependence on the Fourier coefficients E_n while no such complicated dependence exists in (2.2.22) and (2.2.23). The transcendental nature of (4.4.38) is a consequence of the nonlinearity of the surface boundary condition (4.1.4) and is hence unavoidable.

The Fourier analysis technique starts with the one-dimensional drain current transport equation

$$I_D = \mu_n W \left(-Q_m \frac{d\psi_s}{dy} + \frac{1}{\beta} \frac{dQ_m}{dy} \right) \quad (6.1.1)$$

and the sheet charge expression for the continuity of normal displacement.

$$Q_m = -C_{ox} (V_{GB} - \psi_s) - \epsilon_s \psi_x \quad (6.1.2)$$

$$\text{where } \psi_s(y) = \psi(0, y) \quad (6.1.3)$$

$$\psi_x(y) = \frac{\partial \psi}{\partial x}(0, y) \quad (6.1.4)$$

The subtle, but crucial decision that has been made at this point is to base the analysis on the three functions $\psi(0, y)$, $Q_m(y)$ and $\partial\psi/\partial x(0, y)$ instead of on $\psi(0, y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0, y)$ as in the iterative function technique. This is permissible because each set of three functions contains the same amount of information. It is desirable from a physical point of view to replace $\phi_n(y)$ with $Q_m(y)$ since the mobile charge density Q_m is a physically real and measurable quantity, while the quasi-Fermi potential ϕ_n obtains its physical reality indirectly through Q_m and is an unmeasurable quantity. But mostly it is mathematically desirable to replace $\phi_n(y)$ with $Q_m(y)$ because then the exponentials that posed such problems with the iterative function technique disappear. Furthermore, there is then a linear relationship between the three functions $\psi(0, y)$, $Q_m(y)$ and $\partial\psi/\partial x(0, y)$ given by (6.1.2).

The current transport equation (6.1.1) and the continuity of normal displacement equation (6.1.2) are combined by differentiating (6.1.2)

$$\frac{dQ_m}{dy} = C_{ox} \frac{d\psi_s}{dy} - \epsilon_s \frac{d\psi_x}{dy} \quad (6.1.5)$$

and eliminating Q_m from (6.1.1) by substituting (6.1.2) and (6.1.5) into (6.1.1). Upon rearranging this gives

$$I_D = \mu_n W C_{ox} g(y) \quad (6.1.6)$$

where

$$g(y) = (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} + (V_{GB} + \frac{1}{\beta}) \frac{d\psi_s}{dy} - \frac{R t_{ox}}{\beta} \frac{d\psi_x}{dy} \quad (6.1.7)$$

One-dimensional current continuity is then enforced by writing

$$\frac{d}{dy} \left(\frac{I_D}{C_{ox} W} \right) = \frac{d}{dy} (\mu_n g) = 0 \quad (6.1.8)$$

Equation (6.1.8) is the general case of current continuity. A simpler, yet still important case, is that of a constant mobility, where (6.1.8) reduces to

$$\frac{dg}{dy} = 0 \quad (6.1.9)$$

A constant mobility is assumed here to permit a less complex analysis. This then provides a basis for the discussion of the nonconstant mobility case in Chapter 8.

Equation (6.1.9) embodies the principles of both current continuity and the continuity of normal displacement. These two

conditions are transformed into Fourier space by applying the following integral transform to (6.1.9).

$$\int_0^L \frac{dg}{dy} \sin\left(\frac{n\pi y}{L}\right) dy = 0 \quad (6.1.10)$$

where $n = 1, 2, 3, \dots$, and where (6.1.10) means that all the coefficients of the Fourier sine series for dg/dy are zero. Rather than try to perform this integral transform directly, which is algebraically quite complicated, (6.1.10) is integrated by parts to give the equivalent condition

$$\int_0^L g(y) \cos\left(\frac{n\pi y}{L}\right) dy = 0 \quad (6.1.11)$$

Substituting the expression for $g(y)$ from (6.1.7) into (6.1.11) and rearranging yields

$$\begin{aligned} & \int_0^L (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} \cos\left(\frac{n\pi y}{L}\right) dy \\ & + \int_0^L \frac{dh}{dy} \cos\left(\frac{n\pi y}{L}\right) dy = 0 \end{aligned} \quad (6.1.12)$$

$$\text{where } h(y) = \left(V_{GB} + \frac{1}{\beta}\right) \psi_s - \frac{R t_{ox}}{\beta} \psi_x \quad (6.1.13)$$

Integrating the second integral in (6.1.12) by parts gives

$$\int_0^L (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} \cos\left(\frac{n\pi y}{L}\right) dy$$

$$+ h(L) (-1)^n - h(0) + \frac{n\pi}{L} \int_0^L h(y) \sin\left(\frac{n\pi y}{L}\right) dy = 0 \quad (6.1.14)$$

Equation (6.1.14) is now rearranged to

$$\begin{aligned} & \frac{L}{2} \frac{2}{L} \int_0^L (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} \cos\left(\frac{n\pi y}{L}\right) dy \\ & + h(L) (-1)^n - h(0) + \frac{n\pi}{2} (v_{GB}' + \frac{1}{\beta}) \frac{2}{L} \int_0^L \psi_s \sin\left(\frac{n\pi y}{L}\right) dy \\ & - \frac{n\pi}{2} \frac{R t_{ox}}{\beta} \frac{2}{L} \int_0^L \psi_x \sin\left(\frac{n\pi y}{L}\right) dy = 0 \end{aligned} \quad (6.1.15)$$

and the following definitions are introduced

$$F_n = \frac{2}{L} \int_0^L [U(0,y) + V(0,y)] \sin\left(\frac{n\pi y}{L}\right) dy \quad (6.1.16)$$

$$G_n = \frac{2}{L} \int_0^L \left[\frac{\partial U}{\partial x}(0,y) + \frac{\partial V}{\partial x}(0,y) \right] \sin\left(\frac{n\pi y}{L}\right) dy \quad (6.1.17)$$

$$H_n = \frac{2}{L} \int_0^L (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} \cos\left(\frac{n\pi y}{L}\right) dy \quad (6.1.18)$$

It is also useful to note that by applying the orthogonality relation (4.4.37) to (4.4.34) and its partial derivative with respect to x , the following two equations result

$$E_n = \frac{2}{L} \int_0^L Z(0,y) \sin\left(\frac{n\pi y}{L}\right) dy \quad (6.1.19)$$

$$-\frac{n\pi}{L} E_n \tanh\left(\frac{n\pi d}{L}\right) = \frac{2}{L} \int_0^L \frac{\partial Z}{\partial x}(0, y) \sin\left(\frac{n\pi y}{L}\right) dy \quad (6.1.20)$$

Substituting equations (6.1.16) to (6.1.20) into (6.1.15) and isolating E_n gives

$$E_n = \frac{\frac{R t_{ox}}{\beta} G_n - (V_{GB}' + \frac{1}{\beta}) F_n + 2 \frac{h(0) - h(L)}{n\pi} \frac{(-1)^n}{L} - \frac{L}{n\pi} H_n}{V_{GB}' + \frac{1}{\beta} + \frac{R t_{ox}}{\beta} \frac{n\pi}{L} \tanh\left(\frac{n\pi d}{L}\right)} \quad (6.1.21)$$

Equation (6.1.21) represents the Fourier transform of (6.1.9), and is a nonlinear algebraic equation because (6.1.9) is a nonlinear ordinary differential equation. The nonlinearity of (6.1.21) is due entirely to the dependence of H_n on the E_n , caused by the presence of ψ_s , ψ_x and $d\psi_s/dy$ in (6.1.18). This nonlinearity is not very severe especially for small V_{DS} , because in this case $d\psi_s/dy$ is approximately zero making the H_n small enough to be almost negligible in (6.1.21).

6.2 Newton and Quasi-Newton Methods

Since the Fourier series involving the E_n are in practice truncated to N terms, there are actually N equations of the form of (6.1.21). These form a system of N nonlinear algebraic equations in the N unknowns, $E_1, E_2 \dots E_N$. To deal with this system it is best to define the n 'th function $f_n(E)$ as

$$f_n(E) = E_n - \frac{\frac{R t_{ox}}{\beta} G_n - (V_{GB}' + \frac{1}{\beta}) F_n + 2 \frac{h(0) - h(L)}{n\pi} (-1)^n - \frac{L}{n\pi} H_n}{V_{GB}' + \frac{1}{\beta} + \frac{R t_{ox}}{\beta} \frac{n\pi}{L} \tanh(\frac{n\pi d}{L})} \quad (6.2.1)$$

that is obtained by rearranging (6.1.21). The system of N nonlinear equations can then be written compactly as

$$\underline{f}(E) = 0 \quad (6.2.2)$$

$$\text{where } \underline{f} = (f_1, f_2, f_3 \dots f_N)^T \quad (6.2.3)$$

$$\underline{E} = (E_1, E_2, E_3 \dots E_N)^T \quad (6.2.4)$$

The superscript T indicates that a transpose is to be taken making both $\underline{f}$ and $\underline{E}$ column vectors. A good method to use for the solution of (6.2.2) is the N -dimensional Newton's method described by

$$\underline{E}' = \underline{E} - J^{-1} \underline{f}(\underline{E}) \quad (6.2.5)$$

where J is the Jacobian matrix defined by

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial E_1} & \dots & \frac{\partial f_1}{\partial E_N} \\ \vdots & & \vdots \\ \frac{\partial f_N}{\partial E_1} & \dots & \frac{\partial f_N}{\partial E_N} \end{bmatrix} \quad (6.2.6)$$

and $\underline{E}'$ is the next vector in the iterative sequence. See Ortega and Rheinboldt [65] for a good discussion of Newton's method and

Zeidler [66] for a thorough treatment of fixed point processes in general.

Although it is possible to calculate the Jacobian matrix for the nonlinear system under consideration here, it is computationally expensive as it requires of the order of N^3 multiplications. For example to calculate the Jacobian for $N = 40$ requires approximately 64,000 multiplications. A good alternative is to calculate the Jacobian only once at the initial E vector and thereafter estimate it at the subsequent E vectors using Broyden's quasi-Newton method [67]. Quasi-Newton methods are techniques by which the Jacobian matrix for a system of N nonlinear equations, or the Hessian matrix for an N -dimensional optimization problem, may be estimated with little computational effort using iterative or updating techniques. According to Dennis and Moré [68], Broyden's method is the only one that is useful for estimating Jacobians.

Broyden's method requires that the inverse of the Jacobian be replaced by the matrix H so (6.2.5) becomes

$$E' = E - H f(E) \quad (6.2.7)$$

$$\text{where } H = J^{-1} \quad (6.2.8)$$

The initial guess at H may be taken as the identity matrix I , but to speed up convergence for the problem under consideration here the true Jacobian is calculated at the initial E vector and inverted. Both calculating the Jacobian and inverting it take of

the order of N^3 multiplications. After the first iteration H is approximated by the updating formula

$$H' = H + \frac{(s - Hy) s^T H}{s^T Hy} \quad (6.2.9)$$

$$\text{where } s = E_k - E_{k-1} \quad (6.2.10)$$

$$Y = f_k - f_{k-1} \quad (6.2.11)$$

The superscript T indicates a transpose and k is the iteration number. Calculating H using (6.2.9) requires of the order of N^2 multiplications and thus it represents a considerable computational saving.

6.3 Calculation of the Fourier Quantities F_n , G_n and H_n

To calculate the functions $f_n(E)$ and proceed with Newton's method it is necessary to calculate F_n , G_n and H_n from (6.1.16), (6.1.17) and (6.1.18). In the case of F_n and G_n , this is done by Fourier analyzing the expressions

$$\begin{aligned} U(0, y) + V(0, y) = & U(0, y) \\ & + \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \end{aligned} \quad (6.3.1)$$

$$\begin{aligned} \frac{\partial U}{\partial x}(0, y) + \frac{\partial V}{\partial x}(0, y) = & -\sqrt{2 K U(0, y)} \\ & + \frac{1}{R t_{ox}} \sum_{i=0}^{\infty} \frac{\cos u_i}{\alpha_i \sinh(u_i L/d)} [A_i \sinh(u_i \frac{L-y}{d}) + B_i \sinh(u_i y/d)] \end{aligned} \quad (6.3.2)$$

that are obtained from (4.4.39) using (4.3.16) and (4.4.14). The function $U(0,y)$ is calculated from equations (4.3.21) to (4.3.27) and the integrals in (6.1.16) and (6.1.17) are evaluated numerically using the high speed form of Patterson's method described by (5.5.40) to (5.5.46). There are of course N F_n and N G_n coefficients so each integral must be evaluated N times, but this operation need only be done once in the iterative procedure since the F_n and G_n do not depend on the E_n .

It is somewhat more complicated to evaluate H_n coefficients. This is because the H_n depend on the E_n and because (6.1.18) is the Fourier analysis of a product of functions. The easiest way to deal with (6.1.18) is to expand both $R \cos \psi_x - \psi_s$ and $d\psi_s/dy$ into cosine series. This permits the orthogonality of the triple cosine integral to be taken advantage of.

It is easy to render $d\psi_s/dy$ as a cosine series since by using (6.1.16) and (6.1.19) ψ_s may be written as the sine series

$$\psi_s = \sum_{n=1}^{\infty} (E_n + F_n) \sin \left(\frac{n\pi y}{L} \right) \quad (6.3.3)$$

and the derivative of a sine series is a cosine series. But care must be exercised in taking this derivative because (6.3.3) is a sine series where $\psi_s(0)$ and $\psi_s(L)$ are nonzero. Employing the theorem concerning such a differentiation from Tolstov's book on Fourier series [69] it is found that

$$\frac{d\psi_s}{dy} = \frac{B_0}{2} + \sum_{n=1}^{\infty} B_n \cos \left(\frac{n\pi y}{L} \right) \quad (6.3.4)$$

• where $B_0 = C_S$ (6.3.5)

$$B_n = \frac{\hbar\pi}{L} (E_n + F_n) - d_S + (c_S + d_S) (-1)^n \quad (6.3.6)$$

$$c_S = \frac{2}{L} [\psi_S(L) - \psi_S(0)] \quad (6.3.7)$$

$$d_S = \frac{2}{L} \psi_S(0) \quad (6.3.8)$$

To write $R t_{ox} \psi_x - \psi_s$ as a cosine series it is necessary to proceed in a roundabout way because (6.1.19) and (6.1.20) imply that $R t_{ox} Z_x - Z_s$ should be written as the following sine series

$$R t_{ox} Z_x - Z_s = - \sum_{n=1}^{\infty} E_n [1 + R t_{ox} \frac{\hbar\pi}{L} \tanh(\frac{\hbar\pi d}{L})] \sin(\frac{n\pi y}{L}) \quad (6.3.9)$$

where the subscripts x and s have the same meaning for Z as for ψ . Thus it is necessary to convert (6.3.9) to a cosine series using the Half-Range Conversion Theorem given below, which is derived using the standard relationships that define half-range Fourier series.

Half-Range Conversion Theorem

$$\text{If } f(y) = \sum_{n=1}^{\infty} B_n \sin(\frac{n\pi y}{L}) \quad (6.3.10)$$

then $f(y)$ can be converted to the cosine series

$$f(y) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos\left(\frac{n\pi y}{L}\right) \quad (6.3.11)$$

$$\text{where } A_n = \frac{1}{\pi} \sum_{k=1}^{\infty} B_k [1 - (-1)^{k+n}] \left(\frac{1}{k+n} + \frac{1}{k-n}\right) \quad (6.3.12)$$

An analogous set of equations exists to convert cosine series to sine series. To minimize the amount of computation required by (6.3.12) it is best to evaluate $[1 - (-1)^{k+n}]$ using

$$1 - (-1)^i = 2 [i - 2 (i/2)] \quad (6.3.13)$$

where the brackets around $(i/2)$ imply a truncation. This reduces the number of operations necessary to calculate $1 - (-1)^i$ from i multiplications to two multiplications and a division. Using the Half-Range Conversion Theorem $R t_{ox} Z_X - Z_S$ may be written as

$$R t_{ox} Z_X - Z_S = \frac{C_0}{2} + \sum_{n=1}^{\infty} C_n \cos\left(\frac{n\pi y}{L}\right) \quad (6.3.14)$$

where

$$C_n = \frac{2}{\pi} \sum_{k=1}^{\infty} \left\{ E_k \left[1 + R t_{ox} \frac{k\pi}{L} \tanh\left(\frac{k\pi d}{L}\right) \right] \right. \\ \left. [k + n - 2 \left(\frac{k+n}{2}\right)] \left(\frac{1}{k+n} + \frac{1}{k-n}\right) (1 - \delta_{kn}) \right\} \quad (6.3.15)$$

Note that (6.3.13) has been substituted into (6.3.12) to produce (6.3.15), and the $k=n$ condition in (6.3.12) has been replaced with a Kronecker delta in (6.3.15).

To deal with the U and V parts of $R t_{ox} \psi_x - \psi_s$ it is written as

$$R t_{ox} \psi_x - \psi_s = R t_{ox} (U_x + V_x + Z_x) - U_s - V_s - Z_s \quad (6.3.16)$$

$$\text{But by (4.4.14) } R t_{ox} V_x - V_s = 0 \quad (6.3.17)$$

and using (4.3.16)

$$R t_{ox} U_x - U_s = - \frac{\sqrt{2 \epsilon_s q N_B U_s}}{C_{ox}} - U_s \quad (6.3.18)$$

Thus

$$R t_{ox} \psi_x - \psi_s = - \frac{\sqrt{2 \epsilon_s q N_B U_s}}{C_{ox}} - U_s + R t_{ox} Z_x - Z_s \quad (6.3.19)$$

So if the D_n are defined as

$$D_n = \frac{2}{L} \int_0^L \left(- \frac{\sqrt{2 \epsilon_s q N_B U_s}}{C_{ox}} - U_s \right) \cos \left(\frac{n\pi y}{L} \right) dy \quad (6.3.20)$$

where the integral in (6.3.20) is evaluated using the high speed version of Patterson's method, then $R t_{ox} \psi_x - \psi_s$ may be expanded into the cosine series

$$R t_{ox} \psi_x - \psi_s = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos \left(\frac{n\pi y}{L} \right) \quad (6.3.21)$$

where $A_n = C_n + D_n$

(6.3.22)

To evaluate the H_n (6.3.4) and (6.3.21) are substituted into (6.1.18) to give

$$H_n = \frac{2}{L} \int_0^L \left[\frac{A_0}{2} + \sum_{k=1}^{\infty} A_k \cos \left(\frac{k\pi y}{L} \right) \right] \left[\frac{B_0}{2} + \sum_{i=1}^{\infty} B_i \cos \left(\frac{i\pi y}{L} \right) \right] \cos \left(\frac{n\pi y}{L} \right) dy \quad (6.3.23)$$

Multiplying the terms in (6.3.23) and interchanging the order of summation and integration yields

$$\begin{aligned} H_n = & \frac{2}{L} \left[\frac{A_0 B_0}{4} \int_0^L \cos \left(\frac{n\pi y}{L} \right) dy + \frac{A_0}{2} \sum_{i=1}^{\infty} B_i \int_0^L \cos \left(\frac{i\pi y}{L} \right) \cos \left(\frac{n\pi y}{L} \right) dy \right. \\ & + \frac{B_0}{2} \sum_{k=1}^{\infty} A_k \int_0^L \cos \left(\frac{k\pi y}{L} \right) \cos \left(\frac{n\pi y}{L} \right) dy \\ & \left. + \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} A_k B_i \int_0^L \cos \left(\frac{k\pi y}{L} \right) \cos \left(\frac{i\pi y}{L} \right) \cos \left(\frac{n\pi y}{L} \right) dy \right] \quad (6.3.24) \end{aligned}$$

The solutions of the four integrals in (6.3.24) are

$$\int_0^L \cos \left(\frac{n\pi y}{L} \right) dy = 0 \quad (6.3.25)$$

$$\int_0^L \cos \left(\frac{i\pi y}{L} \right) \cos \left(\frac{n\pi y}{L} \right) dy = \frac{L}{2} \delta_{in} \quad (6.3.26)$$

$$\int_0^L \cos \left(\frac{k\pi y}{L} \right) \cos \left(\frac{n\pi y}{L} \right) dy = \frac{L}{2} \delta_{kn} \quad (6.3.27)$$

$$\int_0^L \cos\left(\frac{k\pi y}{L}\right) \cos\left(\frac{i\pi y}{L}\right) \cos\left(\frac{n\pi y}{L}\right) dy = \frac{L}{4} [\delta_{i, (k+n)} + \delta_{i, |k-n|} (1 - \delta_{kn})] \quad (6.3.28)$$

Substituting (6.3.25) to (6.3.28) into (6.3.24), using the Kronecker deltas to eliminate all but one of the summation signs, and rearranging gives the final equation for the H_n :

$$H_n = \frac{1}{2} [A_0 B_n + \sum_{k=1}^{\infty} A_k (B_{k+n} + B_{|k-n|})] \quad (6.3.29)$$

For $n = 0$ (6.3.29) reduces to

$$H_0 = \frac{A_0 B_0}{2} + \sum_{k=1}^{\infty} A_k B_k \quad (6.3.30)$$

Although H_0 is not needed explicitly in the calculation of the E_n , because E_0 is known to be zero, it is a truly fundamental quantity because the drain current may be calculated from it. The drain current expression is obtained by integrating (6.1.6) from the source to the drain end of the channel, and invoking current continuity by removing I_D from the integral because it is constant. Thus in the case of constant mobility under consideration here

$$I_D \int_0^L dy = \mu_n W C_{ox} \left[\int_0^L (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} dy + (V_{GB} + 1/\beta) \int_{\psi_{s0}}^{\psi_{sL}} d\psi_s - \frac{R t_{ox}}{\beta} \int_{\psi_{x0}}^{\psi_{xL}} d\psi_x \right] \quad (6.3.31)$$

Using the fact that by the definition of H_n (6.1.18)

$$H_0 = \frac{2}{L} \int_0^L (R t_{ox} \psi_x - \psi_s) \frac{d\psi_s}{dy} dy \quad (6.3.32)$$

(6.3.31) reduces to

$$I_D = \mu_n \frac{W}{L} C_{ox} \left[\frac{L}{2} H_0 + (V_{GB}' + 1/\beta) (\psi_{sL} - \psi_{s0}) - \frac{R t_{ox}}{\beta} (\psi_{xL} - \psi_{x0}) \right] \quad (6.3.33)$$

This is a most remarkable result because it is a two-dimensional sheet charge expression for the drain current in closed form. Thus it is the two-dimensional analogue of the one-dimensional sheet charge expression derived by Brews [4]. This is a vast improvement over the method of calculating the drain current that was used with the iterative function technique (see equation (5.2.30)). It is also perhaps the basis for a new generation of high speed models of the short-channel MOSFET that could be used for circuit simulation, because this would be the result if simple approximations for H_0 could be found. To use (6.3.33) both the boundary potentials and normal fields must be supplied. The boundary potentials are calculated by the method recommended in Chapter 4 using (4.2.10) to (4.2.15). The boundary normal fields are found by determining the surface derivative of the single/double parabola boundary conditions from (4.2.23) to (4.2.28). Thus

$$\text{If } V_{GB}' > V_{GBCrS} \quad \psi_{x0} = - \frac{2 \psi_{s0}}{x_S} \quad (6.3.34)$$

$$\text{If } V_{GB} \leq V_{GBCrS} \quad \psi_{x0} = c_1 \quad (6.3.35)$$

$$\text{If } V_{GB} > V_{GBCrD} \quad \psi_{xL} = - \frac{2 \psi_{sL}}{x_D} \quad (6.3.36)$$

$$\text{If } V_{GB} < V_{GBCrD} \quad \psi_{xL} = d_1 \quad (6.3.37)$$

6.4 Details of the Iterative Procedure

As mentioned previously, to start the Newton's method iteration it is best to calculate and invert the Jacobian at the initial guess for the E vector. To calculate the individual elements of the Jacobian (6.2.1) is differentiated with respect to E_k :

$$\frac{\partial f_n}{\partial E_k} = \delta_{nk} + \frac{\frac{\partial H_n}{\partial E_k}}{\frac{n\pi}{L} [V_{GB} + 1/\beta + \frac{R t_{ox}}{\beta} \frac{n\pi}{L} \tanh(\frac{n\pi d}{L})]} \quad (6.4.1)$$

and $\partial H_n / \partial E_k$ is obtained by differentiating (6.3.29):

$$\begin{aligned} \frac{\partial H_n}{\partial E_k} = & \frac{1}{2} \left\{ B_n \frac{\partial A_0}{\partial E_k} + A_0 \frac{\partial B_n}{\partial E_k} \right. \\ & \left. + \sum_{i=1}^{\infty} [(B_{i+n} + B_{|i-n|}) \frac{\partial A_i}{\partial E_k} + A_i (\frac{\partial B_{i+n}}{\partial E_k} + \frac{\partial B_{|i-n|}}{\partial E_k})] \right\} \quad (6.4.2) \end{aligned}$$

By differentiating (6.3.6) it is found that

$$\frac{\partial B_i}{\partial E_k} = \frac{i\pi}{L} \delta_{ik} \quad (6.4.3)$$

which reduces (6.4.2) to

$$\begin{aligned} \frac{\partial H_n}{\partial E_k} = & \frac{1}{2} [B_n \frac{\partial A_0}{\partial E_k} + \frac{k\pi}{L} (A_{k+n} + A_{|k-n|}) \\ & + \sum_{i=1}^{\infty} (B_{i+n} + B_{|i-n|}) \frac{\partial A_i}{\partial E_k}] \end{aligned} \quad (6.4.4)$$

The derivative $\partial A_i / \partial E_k$ is calculated by noting that since the D_n are not a function of the E_n then by (6.3.22)

$$\frac{\partial A_i}{\partial E_k} = \frac{\partial C_i}{\partial E_k} \quad (6.4.5)$$

and hence by differentiating (6.3.15) it is found that

$$\begin{aligned} \frac{\partial A_i}{\partial E_k} = & -\frac{2}{\pi} [1 + R t_{ox} \frac{k\pi}{L} \tanh(\frac{k\pi d}{L})] [i + k - 2(\frac{i+k}{2})] \\ & \cdot (\frac{1}{k+i} + \frac{1}{k-i}) (1 - \delta_{ki}) \end{aligned} \quad (6.4.6)$$

The elements of the Jacobian matrix $\partial f_n / \partial E_k$ are determined by calculating the $\partial A_i / \partial E_k$ from (6.4.6), substituting these values into (6.4.4) so that the $\partial H_n / \partial E_k$ are found, and then substituting these values into (6.4.1).

Once the Jacobian matrix is determined it must be inverted to find the matrix H which is done easily and efficiently by Gaussian elimination. This technique involves performing identical elementary row operations on the Jacobian matrix and an

identity matrix of the same dimensions until the Jacobian is converted to an identity matrix [59]. When this point is reached the original identity matrix has now become the inverse of the Jacobian H . The elementary row operations necessary to reduce an arbitrary matrix to the identity matrix break down into three types: forward substitution that makes all elements below the main diagonal zero; normalization of all the elements on the main diagonal to 1; and backwards substitution that makes all elements above the main diagonal zero. In addition, pivoting may be required during forward substitution to avoid numerical inaccuracies. This is the interchange of rows so that no element below the main diagonal has a larger magnitude than the element directly above it on the main diagonal. It was found that pivoting is never necessary for the inversion of the Jacobian because the elements on its main diagonal always have a larger magnitude than any element directly below them. This fact was used to simplify the inversion algorithm and permitted the creation of a fast program to invert the Jacobian matrix.

The initial E vector at which the Jacobian is calculated is obtained by starting with a pre-initial guess of the 0 vector for E . This causes the C_n to be zero by (6.3.15) and hence the A_n are equal to the D_n by (6.3.22) (see Figure 6.4.1 for a flowchart of the algorithm). Upon calculating the B_n from (6.3.6) the H_n may be found from (6.3.29) and then substituted into (6.1.21) to produce the E_n , and hence the initial E vector. This is the E vector at which the Jacobian is calculated and at which the

Figure 6.4.1. Flowchart of the Fourier analysis technique when the mobility is constant.

Newton iteration begins. The iteration ~~then~~ proceeds by calculating the A_n from (6.3.15) and (6.3.22), the B_n from (6.3.6), the H_n from (6.3.29), and the functions $f_n(E)$ from (6.2.1). A new E vector is then produced from the old E vector using (6.2.7), where on the first iteration the H matrix is the inverse of the Jacobian and thereafter it is calculated from Broyden's updating formula (6.2.9). This algorithm converges to the solution for the E vector in typically 2 to 8 iterations. The exceptions to this rule are certain pathological cases that occur when the device is biased just above threshold where as many as 22 iterations may be required. Behavior of this type is consistent with the calculation of boundary potentials using equations (4.3.38) to (4.3.45) where convergence is also slow just above threshold.

The reason for the large number of iterations in certain cases is the widely recognized fault of Newton's method that it tends to overshoot the solution in the early stages of the iteration. This slows convergence and may even produce divergence if the E vector ends up outside the region of convergence that lies in the neighborhood of the solution. The remedy for this problem is to damp Newton's method by limiting the step size on a given iteration. This is done by adding a factor of γ to (6.2.7) to give

$$\underline{E}' = \underline{E} - \gamma \underline{H} \underline{f}(\underline{E}) \quad (6.4.7)$$

It was found that a satisfactory damping method is to simply require that the magnitude of the step vector never exceed the present magnitude of E , and if it does to limit it to the magnitude of E . Hence γ is given by the rules

$$\text{If } |H f(E)| < |E| \text{ then } \gamma = 1 \quad (6.4.8)$$

$$\text{If } |H f(E)| > |E| \text{ then } \gamma = \frac{|E|}{|H f(E)|} \quad (6.4.9)$$

Using this definition for γ damping is generally only performed for pathological cases and even then it is only for the first 1 to 3 iterations. The result is that for these cases the total number of iterations is reduced by as many as 10. In addition, divergence becomes much less likely.

6.5 Performance of the Fourier Analysis Technique

The progress of the iteration may be illustrated by plotting drain current versus iteration number, as was done for the iterative function technique. Using (6.3.33) to evaluate the drain current after each iteration, plots of drain current versus iteration number were produced for one case below threshold, another just above threshold, and a third well above threshold. These plots are displayed in Figures 6.5.1, 6.5.2 and 6.5.3 respectively. The subthreshold plot in Figure 6.5.1 appears to take only 3 iterations to converge but in reality it takes 8 iterations to meet the convergence criterion of I_D changing by less than 1 part in 10^4 from iteration to iteration. The reason for this is that the linear current scale makes the change in I_D

Figure 6.5.1. Drain current versus iteration number for Device 1 biased below threshold where $V_{GS} = 0$, $V_{DS} = 5$ V and $V_{SB} = 0$. See Table 4.3.1 for the device parameters.

Figure 6.5.2. Drain current versus iteration number for Device 1 biased just above threshold where $V_{GS} = 1.2$ V, $V_{DS} = 0$ and $V_{SB} = 0$. See Table 4.3.1 for the device parameters.

Figure 6.5.3. Drain current versus iteration number for Device 1 biased well above threshold where $V_{GS} = 5$ V, $V_{DS} = 5$ V and $V_{SB} = 0$. See Table 4.3.1 for the device parameters.

for the last few iterations unobservable. The plot for operation just above threshold in Figure 6.5.2 shows how pathological such a case is because 22 iterations are required for convergence. Note that I_D does not change very much in the first two or three iterations because the damping restrains the iteration. The case of operation well above threshold in Figure 6.5.3 shows the undamped Newton method at work on an easily handled case. The convergence is rapid as only 3 iterations are required, and without oscillation. Note that except for the pathological case just above threshold, the iterations converge in a monotonic fashion rather than the damped oscillation fashion of the iterative function technique. The difference is that the Fourier analysis technique has made the problem much better behaved and the Newton method accelerates the convergence of the iteration.

Using the Fourier analysis technique as it has been described in this chapter, the E vector is solved for in generally 2 to 8 iterations, and this requires approximately 0.2 to 0.35 seconds of CPU time on an IBM 3081KX computer. The CPU time requirements of the lower level of MINIMOS, where current continuity is treated as being one-dimensional, are between 2 and 60 seconds. Thus when the analytical model uses the Fourier analysis technique it is between 5 and 300 times faster than MINIMOS. This is an accurate comparison because both models solve a two-dimensional Poisson's equation and a one-dimensional current continuity equation for the short-channel MOSFET. Therefore the Fourier analysis technique has produced an

analytical model of the short-channel MOSFET that meets the objective stated in Chapter 1 of creating an analytical model that is significantly faster than comparable numerical models. It has done this by taking advantage of the analytical formulation of the solution for the two-dimensional potential specified by (4.4.39) and making maximum use of analytical manipulations. The other reason for its success is that the only numerical manipulations that are performed (i.e. Newton's method) are done in Fourier space where the problem is simpler because it is cast in terms of algebraic equations rather than differential equations. So clearly the analytical model that makes use of the Fourier analysis technique is fast enough and has sufficient numerical stability. However, to prove that it could really be useful it must be shown that it is also accurate. This is considered in the next chapter.

CHAPTER 7

VERIFICATION OF THE GENERAL MODEL

Just as for the subthreshold model, the accuracy of the general analytical model may be verified by comparing it to measurements or the results of numerical simulations. The choice is determined by the fact that the present version of the general model assumes a uniformly doped substrate and a constant mobility. This makes it necessary to compare the general model to a numerical model because it is then possible to make sure that these assumptions hold for both models. The particular numerical model that was used was a modified version of MINIMOS, where the usual field dependent mobility was discarded and replaced by a constant mobility.

The verification of the general model was done by comparing analytically and numerically generated I_D versus V_{DS} and I_D versus V_{GS} characteristics for two different devices. Device 1 is typical of an approximately 2 μm (gate length) process and Device 2 is typical of a submicron process. Both are n-channel devices and both experience significant short-channel effects. The parameters of these two devices are given in Table 7.1.

Table 7.1. Parameters for Devices 1 and 2. (This table is identical to Table 4.3.1 and is repeated here for convenience.)

Device Type	L	W	N_B	N_S	t_{ox}	r_j	V_{FB}
	(μm)	(μm)	(cm^{-3})	(cm^{-3})	($\text{\AA}$)	(μm)	(V)
Device 1	1.0	24	2×10^{16}	2.3×10^{20}	500	0.532	-0.72
Device 2	0.5	24	4×10^{16}	2.0×10^{20}	150	0.2	-0.207

It is clear from Table 7.1 that Device 2 is a scaled down version of Device 1 because, although its channel length has been reduced by a factor of 2, its substrate doping has been increased and oxide thickness reduced to counterbalance this change. As is seen in the figures below, the parameters that have not been scaled in a similar manner are the bias voltages which are the same for both devices. This should be born in mind while comparing the results for the two devices as it makes the short-channel effects more pronounced for Device 2 than Device 1. The flatband voltages listed in Table 7.1 should be interpreted as the true flatband voltage modified by a surface implant that is shallow enough so as to leave the rest of the substrate essentially uniform. The dose of the surface implant, or in other words the magnitude of V_{FB} , has been chosen to give each device a long-channel threshold voltage of 1 V.

The comparison of analytically and numerically generated I_D versus I_{DS} characteristics for Devices 1 and 2 is shown in Figures 7.1, 7.2 and 7.3. The analytical characteristics were calculated, as in Chapter 6, from the general solution of Poisson's equation and the current continuity equation, using the Fourier analysis technique and the closed form expression for the drain current (see equation (6.3.33)). The characteristics are in good agreement for both devices under all the bias conditions shown. No fitting parameters were employed to obtain this agreement. This shows that the general analytical model is accurate in the triode and saturation regions. The only anomalies present in these results are some ripples in the analytical characteristics, which are especially evident in the saturation regions of the $V_{GS} = 4 \text{ V}$ and $V_{GS} = 5 \text{ V}$ characteristics. It is probable that the ripples are caused by numerical noise generated by the iterative procedures used with the analytical model. It has not been investigated, but it is virtually certain, that the ripples could be eliminated by tightening the convergence criteria of the iterative procedures. It is worth noting that careful examination of the MINIMOS characteristics in Figures 7.1, 7.2 and 7.3 show that they also exhibit ripples, and in fact have more numerical noise than is present in the analytical characteristics, indicating that the problem is of a general nature.

Figure 7.1. I_D versus V_{DS} for Device 1 where $V_{SB} = 0$.

Figure 7.2. I_D versus V_{DS} for Device 2 where $V_{SB} = 0$.

Figure 7.3, I_D versus V_{DS} for Device 1 where $V_{SB} = 2V$.

The analytically and numerically generated I_D versus V_{GS} characteristics are compared in Figures 7.4 and 7.5. The third type of characteristic labeled as "Gummel" is discussed shortly. It is seen that good agreement has been obtained between the characteristics calculated by the closed form drain current equation and by MINIMOS. The agreement is somewhat worse for the subthreshold part of the $V_{DS} = 5$ V curve, which is probably due to inaccuracies in the double parabola boundary condition used on the drain side of the rectangular box, and the use of the gradual channel approximation in the derivation of the long-channel trial function $U(x,y)$. Despite this, it is reasonable to conclude that the general analytical model has fairly good accuracy in subthreshold.

Round-off error prevents the closed form equation from reaching into deep subthreshold, and this is what produces the spurious tails on the closed form equation characteristics. This same problem is encountered by the one-dimensional sheet charge expression derived by Brews [4] that is directly related to the two-dimensional closed form equation used here. MINIMOS also encounters difficulty in calculating drain currents in deep subthreshold as is demonstrated by the scatter of its points in Figures 7.4 and 7.5. The way to overcome the difficulties caused by round-off error in deep subthreshold is to calculate the drain current from the expression that results from Gummel's algorithm (see equation (2.4.14)). This expression has already been used with the subthreshold model and permits the calculation of the

Figure 7.4. I_D versus V_{GS} for Device 1.

Figure 7.5. I_D versus V_{GS} for Device 2.

drain current arbitrarily far into subthreshold as is shown in Figures 7.4 and 7.5. Note that there is good agreement between the closed form and Gummel expressions in the region where they overlap indicating that they are equivalent. The reason Gummel's expression was not used to calculate a characteristic for $V_{DS} = 5$ V, is because the large drain voltage necessitates the calculation of the field reversal point so that the shortened channel length L' can be calculated (see equation (2.4.18)), and the equations to do this have not yet been derived for the general model.

It has been shown in this chapter that the general analytical model is accurate in the subthreshold, triode and saturation regimes. This demonstrates that the general solution for the two-dimensional potential is valid, the Fourier analysis technique is effective, and that the closed form drain current equation is accurate in all the normal regimes of operation. It is thus concluded that the general two-dimensional analytical model is a major step forward in short-channel MOSFET modelling because it is fast, accurate and numerically stable.

4

of/de

4

Micro-D

CHAPTER 8

PARTIALLY COMPLETED AND FUTURE RESEARCH

8.1 Introduction

This chapter has the dual purpose of presenting partially completed additional research, and indicating the many avenues of future research that stem from the work of the thesis. The partially completed research consists of material that significantly extends the work presented so far, but which has not been implemented or tested on a computer. The material relates to: the subthreshold model; the general model; and the methods used to solve for the E vector.

The most important of these is the additional research on the general model. As was indicated in Chapter 1, the aim of the research undertaken within the scope of the Ph.D. thesis, was to derive a general analytical model for the short-channel MOSFET for the case of a uniformly doped substrate and a constant mobility, but in such a way that the method of solution could be extended to the general case of a nonuniform substrate doping profile and a field dependent mobility. The extension of the solution method to the general case forms a major part of the additional research presented in this chapter and is covered in Section 8.3. The additional research that relates to the subthreshold model and the methods used to solve for the E vector is presented in Sections 8.2 and 8.4 respectively. Sections 8.2, 8.3 and 8.4 also deal with possibilities for future research

related to the subthreshold model, the general model, and the methods used to solve for the E vector respectively. The chapter concludes with Section 8.5 that deals with two possibilities for future research which are not related to these three topics.

8.2 Improvements to the Subthreshold Model

8.2a The Side Boundary Conditions

An obvious improvement that could be made to the subthreshold model, is to replace the cubic side boundary conditions used in the subthreshold Poisson solution, with the highly accurate single/double parabola side boundary conditions developed for use with the general model. This would increase the accuracy of the subthreshold solution of Poisson's equation, the subthreshold current model, and the short-channel threshold voltage model. It is evident that substantial improvement in the drain side boundary condition $\psi(x,L)$ is possible if it is examined when V_{DS} is large. A comparison is made between the cubic and single/double parabola side boundary conditions for a V_{DS} of 5 V in Figures 8.2.1 and 8.2.2 using devices whose parameters are stated on the figures. The cubic boundary conditions give only a rough estimation of the true potential because the surface potential has been set too high. The single/double parabola boundary conditions are much more accurate because for them the surface potential is not fixed but determined by the continuity of normal displacement.

(a)

(b)

Figure 8.2.1. (a) Comparison of the cubic drain side boundary condition and (b) the single/double parabola drain side boundary condition to the results of MINIMOS simulations for a device whose parameters are $L = 3 \mu\text{m}$, $W = 24 \mu\text{m}$, $N_B = 5 \times 10^{15} \text{ cm}^{-3}$, $N_S = 2.3 \times 10^{20} \text{ cm}^{-3}$, $t_{ox} = 500 \text{ \AA}$, $r_j = 0.5 \mu\text{m}$, $V_{FB} = -0.50 \text{ V}$ and whose applied voltages are $V_{GS} = 0.5 \text{ V}$, $V_{DS} = 5 \text{ V}$, $V_{SB} = 0$.

(a)

(b)

Figure 8.2.2. (a) Comparison of the cubic drain side boundary condition and (b) the single/double parabola drain side boundary condition to the results of MINIMOS simulations for a device whose parameters are $L = 1 \mu\text{m}$, $W = 24 \mu\text{m}$, $N_B = 2 \times 10^{16} \text{ cm}^{-3}$, $N_S = 2.3 \times 10^{20} \text{ cm}^{-3}$, $t_{\text{ox}} = 500 \text{ \AA}$, $r_j = 0.532 \mu\text{m}$, $V_{\text{FB}} = -0.72 \text{ V}$ and whose applied voltages are $V_{\text{GS}} = 0.5 \text{ V}$, $V_{\text{DS}} = 5 \text{ V}$, $V_{\text{SB}} = 0$.

To make complete use of the single/double parabola boundary conditions with the subthreshold model they would need to be modified. The nonuniformity trial function $T(x)$ would have to be included in their formulation so that they could be used with nonuniformly doped substrates. The manner in which this is done should be guided by the parabolic boundary conditions (2.1.17) and (2.1.18) that have been modified for use with a nonuniform substrate. Thus $T(x)$ should be added to the single/double parabola boundary conditions and $T(0)$ should be subtracted from appropriate parts of the expressions so that the surface potential is not disturbed.

The single/double parabola boundary conditions would also have to be modified for use with the threshold voltage model. This is because they have a transcendental dependence on V_{GS} resulting in a complicated transcendental expression for the threshold voltage. To alleviate this problem the mobile charge at points $(0,0)$ and $(0,L)$ should be neglected because the device is in subthreshold, making the dependence on V_{GS} quadratic. This would then permit the expression for the threshold voltage to be closed form.

8.2b Defining Threshold

Another unsolved problem is that defining threshold in terms of a minimum surface potential of $2\phi_p + V_{SB}$ is not adequate for short-channel devices (see Section 3.1). A method to investigate what condition is applicable for short-channel devices is to

determine threshold voltages from calculated results in the same way they are found from measurements. This is done by using the general model to produce I_D versus V_{GS} plots and then finding the threshold voltage by extrapolating the region of greatest slope to the V_{GS} axis. The minimum surface potential at threshold can then be determined from the two-dimensional potential. This method could also be used with a numerical model like MINIMOS although numerical scatter could pose a problem in extrapolating the I_D versus V_{GS} plot. By determining the minimum surface potential at threshold for a variety of device structures and biases it should be possible to determine a more meaningful threshold condition than is afforded by using $2\phi_p + V_{SB}$.

8.3 Improvements to the General Model

8.3a Including a Field Dependent Mobility

The analytical solution of a two-dimensional Poisson's equation and a one-dimensional current continuity equation for the short-channel MOSFET that was presented in Chapters 4 and 6, is for the special case of a uniformly doped substrate and a constant mobility. As discussed in Chapter 1, it can be used as the basis of a more general model that includes the effects of a nonuniform substrate doping profile and an electric field dependent mobility. Showing how these effects may be included is the purpose of this section. Once these extensions are added to the model it will be capable of simulating practical small geometry devices; however it will of necessity be more

mathematically complex. A field dependent mobility is included by modifying the Fourier analysis technique for a constant mobility. This is achieved by recalling equation (6.1.6):

$$I_D = \mu_n W C_{ox} g(y) \quad (8.3.1)$$

and rearranging and differentiating it to

$$\frac{d}{dy} \left(\frac{I_D}{\mu_n W C_{ox}} \right) = \frac{dg}{dy} \quad (8.3.2)$$

Current continuity is then applied by removing I_D from the derivative. Thus

$$\frac{dg}{dy} - \frac{I_D}{W C_{ox}} \frac{d}{dy} \left(\frac{1}{\mu_n} \right) = 0 \quad (8.3.3)$$

Transforming to Fourier space,

$$\int_0^L \frac{dg}{dy} \sin \left(\frac{n\pi y}{L} \right) dy - \frac{I_D}{W C_{ox}} \int_0^L \frac{d}{dy} \left(\frac{1}{\mu_n} \right) \sin \left(\frac{n\pi y}{L} \right) dy = 0 \quad (8.3.4)$$

integrating once by parts and defining

$$P_n = \frac{2}{L} \int_0^L \frac{1}{\mu_n} \cos \left(\frac{n\pi y}{L} \right) dy \quad (8.3.5)$$

yields

$$\int_0^L g(y) \cos \left(\frac{n\pi y}{L} \right) dy - \frac{L I_D P_n}{2 W C_{ox}} = 0 \quad (8.3.6)$$

Completing the transformation of $g(y)$ into Fourier space, as was done in Chapter 6, and then rearranging, allows the functions for $f_n(E)$ to be defined as

$$f_n(E) = E_n$$

$$= \frac{\frac{R t_{ox}}{\beta} G_n - (V_{GB}' + 1/\beta) F_n + 2 \frac{h(0) - h(L) (-1)^n}{n\pi} - \frac{L}{n\pi} H_n + \frac{L I_D P_n}{n\pi W C_{ox}}}{V_{GB}' + 1/\beta + \frac{R t_{ox}}{\beta} \frac{n\pi}{L} \tanh\left(\frac{n\pi d}{L}\right)} \quad (8.3.7)$$

So the system of N nonlinear equations in N unknowns that must be solved to calculate the E_n is of the same form as it was for a constant mobility except for the term $L I_D P_n / (n\pi W C_{ox})$ (see equation (6.2.1)).

The drain current for a field dependent mobility is calculated in a similar manner to the constant mobility case by integrating the current transport equation, with the exception that the mobility is placed inside the integral:

$$I_D \frac{1}{L} \int_0^L \frac{dy}{u_n} = \frac{W}{L} C_{ox} \left[\int_0^L (R t_{ox} \psi_x - \psi_s) \cdot \frac{d\psi_s}{dy} dy + (V_{GB}' + 1/\beta) \int_{\psi_{s0}}^{\psi_{sL}} d\psi_s - \frac{R t_{ox}}{\beta} \int_{\psi_{x0}}^{\psi_{xL}} d\psi_x \right] \quad (8.3.8)$$

Defining the averaged mobility as

$$u_n^* \triangleq \left[\frac{1}{L} \int_0^L \frac{dy}{u_n} \right]^{-1} \quad (8.3.9)$$

allows (8.3.8) to be reduced to

$$I_D = \mu_n^* \frac{W}{L} C_{ox} \left[\frac{L}{2} H_0 + (V_{GB} + 1/\beta) (\psi_{sL} - \psi_{s0}) - \frac{R t_{ox}}{\beta} (\psi_{xL} - \psi_{x0}) \right] \quad (8.3.10)$$

which is of the same form as the drain current equation for a constant mobility (6.3.33).

To solve for the E_n for a field dependent mobility, it is necessary to begin by estimating $\mu_n(y)$ and I_D and then calculate the P_n from (8.3.5) (see Figure 8.3.1 for a flowchart of the algorithm). The initial guess for the E vector is constructed by making use of the P_n and the same method as for the constant mobility algorithm. The functions $f_n(E)$ are now calculated and a single Newton iteration performed yielding a new E vector from which $\mu_n(y)$ and I_D are determined. The Newton method then proceeds as for the constant mobility case until convergence is obtained. The initial calculation of the Jacobian should probably be done using exactly the same equations as for the constant mobility case. This is because while the dependence of the new term $L I_D P_n / (n \mu W C_{ox})$ in (8.3.7) is very complicated, it is not the dominant dependence, and so can safely be ignored. Thus the only changes required to include a field dependent mobility in the Poisson solution are that initial guesses for I_D and $\mu_n(y)$ must be supplied and the nonlinear equations in the E_n are slightly more complicated.

Figure 8.3.4. Flowchart of the Fourier analysis technique when the mobility is field dependent.

8.3b Including a Nonuniform Substrate Doping Profile

The other major extension to the general model is to give it the capability to deal with devices that have nonuniformly doped substrates. To derive such a model it is essential to obtain the trial functions $T(x,y)$ and $U(x,y)$ necessary to achieve Laplace reduction. The physical meaning of these functions is that the sum $T(x,y) + U(x,y)$ is the two-dimensional potential for a MOSFET in the long-channel limit. The general case is similar to the subthreshold case in that both require the trial function to be split into a uniformity trial function $U(x,y)$ and a nonuniformity trial function $T(x,y)$. This permits $T(x,y)$ to be used to modify the side boundary conditions so that the Fourier analysis needed to solve the Laplace problem for $V(x,y)$ is simple (see Section 2.2).

For the general case where the MOSFET may be in the subthreshold, triode or saturation regions, and the substrate may be nonuniformly doped in the x direction the two-dimensional Poisson's equation is given by

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{q}{\epsilon_s} [N_B + N_n(x)] = K + f(x) \quad (8.3.11)$$

and is broken into four parts:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = K \quad (8.3.12)$$

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x) \quad (8.3.13)$$

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \quad (8.3.14)$$

$$\frac{\partial^2 Z}{\partial x^2} + \frac{\partial^2 Z}{\partial y^2} = 0 \quad (8.3.15)$$

where $T(x,y)$ and $U(x,y)$ are the trial functions and $V(x,y)$ and $Z(x,y)$ are the solutions to Laplace Problems 1 and 2 respectively (see equations (4.4.11) to (4.4.20)). The significance of the last two functions is that $V(x,y)$ includes the effect of the side boundary conditions, and $Z(x,y)$ includes the effect of the mobile charge induced by short-channel effects. It is convenient to define $S(x,y)$ to be the sum of the trial functions. Thus

$$S(x,y) = T(x,y) + U(x,y) \quad (8.3.16)$$

and hence $S(x,y)$ is the two-dimensional potential of the MOSFET in the long-channel limit which is given by a simultaneous solution of a one-dimensional Poisson's equation and a one-dimensional current continuity equation.

To obtain the one-dimensional Poisson's equation the y derivatives in (8.3.12) and (8.3.13) are neglected in accordance with the gradual channel approximation to give

$$\frac{\partial^2 U}{\partial x^2} = K \quad (8.3.17)$$

$$\frac{\partial^2 T}{\partial x^2} = f(x) \quad (8.3.18)$$

Solving these two differential equations and applying the zero potential and zero electric field at the depletion depth $x_d(y)$ (which neglects the small field due to the nonuniform doping) yields

$$U(x,y) = \frac{K}{2} [x_d(y) - x]^2 \quad (8.3.19)$$

$$T(x,y) = G(x) - G(x_d(y)) + [x_d(y) - x] F(x_d(y)) \quad (8.3.20)$$

where $F(x)$ and $G(x)$ are the first and second integrals of $f(x)$ (see equations (2.2.8) and (2.2.9)). Summing (8.3.19) and (8.3.20) the two-dimensional potential $S(x,y)$ is obtained as

$$S(x,y) = \frac{K}{2} [x_d(y) - x]^2 + G(x) - G(x_d(y)) + [x_d(y) - x] F(x_d(y)) \quad (8.3.21)$$

An important concept needed to determine $S(x,y)$ is that the long-channel depletion depth $x_d(y)$ may be determined from the surface potential $S_s = S(0,y)$ and vice-versa. To determine S_s from x_d (8.3.21) is simply evaluated at $x = 0$.

$$S_s = \frac{K}{2} x_d^2 + G(0) - G(x_d) + x_d F(x_d) \quad (8.3.22)$$

To calculate x_d from S_s (8.3.22) is differentiated and the following Newton-Raphson iterative equation constructed:

$$x_d = x_d - \frac{K/2 x_d^2 + G(0) - G(x_d) + x_d F(x_d) - S_s}{x_d [K + f(x_d)]} \quad (8.3.23)$$

where a good initial guess for (8.3.23) is the uniform substrate depletion depth obtained from (4.3.11).

$$x_d = \sqrt{\frac{2}{K} S_s} \quad (8.3.24)$$

Note that S_s is a closed form function of x_d but x_d is a transcendental function of S_s . Thus if there is a choice between doing one calculation or the other, calculating S_s as a function of x_d should be chosen.

Another important equation is the sheet charge expression for the mobile charge density Q_m for a nonuniformly doped substrate. This is obtained from the following statement of the continuity of normal displacement for the long-channel trial function

$$Q_m = -C_{ox} (V_{GB} - S_s) + q \int_0^{x_d} [N_B + N_n(x)] dx \quad (8.3.25)$$

as

$$Q_m = -C_{ox} (V_{GB} - S_s) + \epsilon_s [K x_d + F(x_d) - F(0)] \quad (8.3.26)$$

Since x_d may be calculated from S_s using (8.3.23), then (8.3.26) gives Q_m as a function of S_s .

Having derived the relationships between Q_m , S_s and x_d , it is now possible to combine the one-dimensional Poisson's equation with the one-dimensional current continuity equation. Note, as this procedure develops, how similar it is to the derivation of the trial function $U(x,y)$ for the general model when the substrate is uniform (see Section 4.3). To begin, the one-dimensional current transport equation

$$I_D = - \mu_n W Q_m \frac{dS_s}{dy} + \frac{\mu_n}{\beta} W \frac{dQ_m}{dy} \quad (8.3.27)$$

is rearranged to

$$\frac{I_D}{\mu_n W} = P = - Q_m \frac{dS_s}{dy} + \frac{1}{\beta} \frac{dQ_m}{dy} \quad (8.3.28)$$

where current continuity has been invoked by setting $I_D/(\mu_n W)$ equal to a constant P . This assumes that for the trial functions the mobility is constant which does not necessarily limit the mobility to being constant for the rest of the solution of $\psi(x,y)$. The reason is that as long as the trial functions achieve Laplace reduction it does not really matter what their physical meaning is, although it is of course helpful if they do have some physical meaning. Thus if a constant mobility is assumed in the derivation of the trial functions then this does not preclude using a field dependent mobility in the calculation of the E vector by the aforementioned method of Section 8.3a.

Integrating (8.3.28) gives

$$Py + Q = - \int Q_m dS_s + \frac{Q_m}{\beta} \quad (8.3.29)$$

If $Q_m(S_s)$ was a function that could be integrated to yield a closed form answer, then it would be possible to proceed as in the uniform substrate case. The fact that it is not is overcome by defining the antiderivative function

$$A(S_s) = \int Q_m(S_s) dS_s \quad (8.3.30)$$

$$\text{Thus } Py + Q = - A(S_s) + \frac{Q_m(S_s)}{\beta} \quad (8.3.31)$$

The two boundary potentials

$$S_{SS} = S_s(0) \quad (8.3.32)$$

$$S_{SD} = S_s(L) \quad (8.3.33)$$

whose calculation will be discussed shortly, are used to calculate the constant P from (8.3.31). This yields

$$P = - \frac{\int_{S_{SS}}^{S_{SD}} Q_m(S_s) dS_s + \frac{Q_m(S_{SD}) - Q_m(S_{SS})}{\beta}}{L} \quad (8.3.34)$$

The constant Q can be formally determined in the same way to be

$$Q = - A(S_{SS}) + \frac{Q_m(S_{SS})}{\beta} \quad (8.3.35)$$

but this is not useful as the antiderivative function $A(S_{SS})$ can only be calculated numerically in the form of a definite integral as in (8.3.34). So instead of calculating Q , (8.3.28) is integrated from 0 to y to give

$$Py + \int_{S_{SS}}^{S_S} Q_m(S_S) dS_S + \frac{Q_m(S_{SS}) - Q_m(S_S)}{\beta} = 0 \quad (8.3.36)$$

Upon differentiating this expression using (8.3.22) and (8.3.26), the following Newton-Raphson iterative equation in S_S is constructed

$$S'_S = S_S - \frac{Py + \int_{S_{SS}}^{S_S} Q_m(S_S) dS_S + \frac{Q_m(S_{SS}) - Q_m(S_S)}{\beta}}{Q_m(S_S) - \frac{1}{\beta} (C_{ox} + x_d^{S_S})} \quad (8.3.37)$$

The procedure to calculate S_S from this equation is to first obtain an initial guess from the constant substrate doping equation (4.3.21), and then iterate using (8.3.37). Upon each iteration it is necessary to calculate x_d in an iterative fashion from (8.3.23) and to numerically integrate $Q_m(S_S)$ from S_{SS} to S_S . The numerical integral is most easily done by substituting (8.3.26) for Q_m , doing as much as possible of the integral analytically, and converting the remainder to an integral over x_d so that the transcendental dependence of x_d on S_S only has to be dealt with in the limits of integration. This yields

$$\int_{S_{SS}}^{S_S} Q_m(S_S) dS_S = \frac{C_{ox}}{2} [(V_{GB}' - S_S)^2 - (V_{GB}' - S_{SS})^2]$$

$$- \epsilon_s F(0) (S_S - S_{SS}) + \epsilon_{s_{x_d(S_{SS})}} \int_{x_d(S_{SS})}^{x_d(S_S)} [Kx_d + F(x_d)] [K + f(x_d)] x_d dx_d$$

(8.3.38)

Once $S_S(y)$ is calculated from (8.3.37), $x_d(y)$ may be calculated from (8.3.23) and this then defines the trial functions $S(x,y)$, $T(x,y)$ and $U(x,y)$ by (8.3.21), (8.3.20) and (8.3.19) respectively.

The only detail that has not yet been dealt with in the calculation of the trial functions is the calculation of the boundary potentials S_{SS} and S_{SD} . These are calculated from the one-dimensional Shockley-Poisson equation

$$\frac{d^2 S}{dx^2} = \frac{1}{2} \frac{dE_x^2}{dS} = \frac{q}{\epsilon_s} [n_i e^{\beta(S-\phi_n)} - n_i e^{\beta(\phi_p-S)} + N(x)] \quad (8.3.39)$$

where $N(x)$ is the substrate doping profile. Integrating (8.3.39) from the depletion edge to the surface and neglecting insignificant terms yields

$$E_s^2 = \frac{2q}{\beta \epsilon_s} [n_i e^{\beta(S_s - \phi_n)} - n_i e^{\beta \phi_p}] + \frac{2q}{\epsilon_s} \int_0^{x_d} N(x) E_x dx \quad (8.3.40)$$

An expression for the quantity $N(x) dx$ is found by summing (8.3.17) and (8.3.18) to give

$$\frac{\partial^2 S}{\partial x^2} = K + f(x) = \frac{q}{\epsilon_s} N(x) dx \quad (8.3.41)$$

and then rewriting (8.3.41) in terms of the electric field component E_x

$$N(x) dx = - \frac{\epsilon_s}{q} dE_x \quad (8.3.42)$$

Substituting (8.3.42) into (8.3.40) and noting that

$$N(x_d) = n_i e^{\beta \phi_p} \quad (8.3.43)$$

it is found that

$$E_s^2 = \frac{2q}{\beta \epsilon_s} [n_i e^{\beta(S_s - \phi_n)} - N(x_d)] + \int_0^{E_x^2(0)} 2 E_x dE_x \quad (8.3.44)$$

Performing the integral and substituting for E_s using

$$E_s = \frac{C_{ox}}{\epsilon_s} (V_{GB} - S_s) \quad (8.3.45)$$

gives

$$\begin{aligned} n_i e^{\beta(S_s - \phi_n)} - N(x_d) + \frac{\beta \epsilon_s}{2q} [K x_d + F(x_d) - F(0)]^2 \\ - \beta C (V_{GB} - S_s)^2 = 0 \end{aligned} \quad (8.3.46)$$

$$\text{where } C = \frac{C_{ox}^2}{2 q \epsilon_s} \quad (8.3.47)$$

Differentiating (8.3.46) with respect to x_d permits the following Newton-Raphson equation in x_d to be created.

$$\begin{aligned} x_d' = x_d - & (n_i e^{\beta(S_s - \phi_n)} - N(x_d) + \frac{\beta \epsilon_s}{2q} [K x_d + F(x_d) - F(0)]^2 \\ & - \beta C (V_{GB} - S_s)^2) / [\beta x_d [K + F(x_d)] (n_i e^{\beta(S_s - \phi_n)} \\ & + \frac{\epsilon_s}{q} [K + \frac{F(x_d) - F(0)}{x_d}] + 2 \cdot C (V_{GB} - S_s)) - \frac{dN}{dx_d}] \quad (8.3.48) \end{aligned}$$

The reason (8.3.48) is written in terms of x_d rather than S_s is that it is computationally most efficient to calculate $x_d(0)$ and $x_d(L)$ and then obtain $S_s(0)$ and $S_s(L)$ (i.e. S_{ss} and S_{sd}) from them. This is because there is a transcendental dependence of x_d on S_s , but there is a closed form dependence of S_s on x_d . The best way to obtain initial guesses for (8.3.48) is to neglect the nonuniformity in the substrate doping and use the uniform substrate initial guesses given by (4.3.43) and (4.3.44). These initial guesses for S_{ss} and S_{sd} are converted to initial guesses for $x_d(0)$ and $x_d(L)$ using (8.3.24).

Once the boundary potentials are determined from (8.3.48) the constants P and Q can be calculated, and the determination of the trial functions $T(x,y)$ and $U(x,y)$ is completed. This permits the reduction of Poisson's equation to Laplace's equation and the

two Laplace problems for $V(x,y)$ and $Z(x,y)$ must then be solved. The solutions for $V(x,y)$ and $Z(x,y)$ would only differ from those described in Chapter 4 in two ways: first, the side boundary conditions for a nonuniform substrate would be different than those for a uniform substrate and this would change the expressions for the coefficients A_i and B_i in the equation for $V(x,y)$. Second, the equations for the Fourier coefficients F_n and G_n (6.1.16) and (6.1.17) would have to include $T(x,y)$. The details of these changes and the rest of the general Poisson solution for a nonuniform substrate have not yet been worked out and must be left for further research.

8.3c Other Improvements

There are other improvements that could be made to the general solution of Poisson's equation irrespective of whether the substrate is uniformly or nonuniformly doped. One such possibility would be to use trial functions that do not employ the gradual channel approximation. In other words the y derivative in equations (4.3.1), (8.3.12) and (8.3.13) would not be neglected. To date, no trial function fitting this description has been identified. But if one was identified it would improve the accuracy of the model because then Poisson's equation could be exactly satisfied. The improvement in accuracy, however, would probably be quite small, because given the agreement between analytically and numerically simulated results found in Chapter 7, and the error incurred by neglecting the y derivative discussed in Chapter 4, it appears that using

the gradual channel approximation to derive the trial functions leads to a virtually insignificant error.

Another possible improvement is to use a two-dimensional solution of Laplace's equation in the gate oxide. At present a one-dimensional Laplace's equation is used and it appears as the term $C_{ox} [V_{GB} - \psi(0,y)]$ in the surface boundary condition (4.1.4). The two-dimensional Laplace's equation would use as its boundary conditions continuity of normal displacement at the silicon-silicon dioxide interface, linear potentials along the sides of the gate oxide rectangle at $y = 0$ and $y = L$ (see Figure 4.1.1), and the constant gate potential along the top of the rectangle. The resulting Laplace problem could be solved by the standard separation of variables method. Clearly a two-dimensional analysis of the potential in the gate oxide will be more accurate than a one-dimensional analysis, but the improvement in accuracy would probably be quite small for thin gate oxides which are the usual case of interest.

Other improvements that could conceivably be made to the general solution are to employ more physically accurate boundary conditions, and to use a two-dimensional current continuity equation. The side and bottom boundary conditions that have been utilized are a compromise between accuracy and mathematical simplicity. Thus it would seem there is room for improvement in their accuracy; however this would inevitably lead to greater mathematical complexity. Employing a two-dimensional current continuity equation would greatly increase the mathematical

complexity because Poisson's equation would then become the highly nonlinear Shockley-Poisson equation given by (2.1.1). The analytical methods required to solve this equation would differ greatly from those outlined in this thesis. Thus any further discussion of the solution must be assigned to further research.

8.4 Solving for the E Vector

One of the three problems that was responsible for the slow convergence of the iterative function technique solution for the E vector presented in Chapter 5, was the presence of numerical noise. This noise is caused by the finite number of terms in the Fourier and Chebyshev series representations of $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$. It is a problem because although the noise level is fairly low, it is amplified by the exponential dependence of $Q_m(y)$ on $\psi(0,y)$ and $\phi_n(y)$ and then fed back into the iterative system.

A method that could be used to reduce this noise problem is to dispense with the Chebyshev series storage of $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$, and simply store the values of these functions at the 200 to 300 points necessary to do the calculations. This would eliminate the noise due to Chebyshev series but not the noise due to Fourier series. The Fourier series noise could be reduced by simply using more terms.

Reducing the noise level should speed up convergence because then there is less randomness to the search for the correct E vector. Storing the functions $\psi(0,y)$, $\phi_n(y)$ and $\partial\psi/\partial x(0,y)$ at only the points they are needed can also be used to speed up

convergence by using the vectors of stored function values in a Newton iteration. A Jacobian matrix could then be calculated using Broyden's quasi-Newton method and used in the iterative search for the correct function vectors. Reducing the noise and using Newton techniques would certainly speed up the iterative function technique, but only by implementing and testing these ideas could it be determined whether they would produce a model that was fast enough to be useful.

The other method of solving for the $\underline{E}$ vector, the Fourier analysis technique, is interesting not only because it produces a fast analytical model of the short-channel MOSFET, but also because it could have much wider applications. It appears that it is related to a class of techniques called "Fast Poisson Solvers" [70]. These techniques use a grid, like the finite difference or finite element techniques, but then transform the problem into the Fourier domain using Fast Fourier Transforms (FFT) and finish the solution of the problem there. The Fast Poisson Solvers are restricted to problems whose domain is geometrically regular such as a rectangle or a circle. The Fourier analysis technique is similar in that it does most of its manipulations in Fourier space and it can only be used on geometrically regular domains, but it is different in that it does not use a grid and it achieves its transformations into Fourier space using the high speed form of Patterson's method (see equations (5.5.40) to (5.5.46)) rather than the Fast Fourier Transform. By analogy to the Fast Poisson Solvers it appears that it might be possible to extend the Fourier analysis

technique to a more general method. This general method would consist of beginning with a separable two-dimensional Fourier series as the form of the solution, then transforming the entire Laplace or Poisson problem into the Fourier domain using Patterson's method and solving the problem there. Solutions of this type might greatly extend the range of application for analytical solutions of Poisson's equation for the MOSFET. No details of this concept have yet been formulated so its implementation must be left for further research.

8.5 New Areas of Research

There are two areas to which the analytical solutions of Poisson's equation described in this thesis could be applied, but that have not been previously mentioned. These are the calculation of the intrinsic capacitances for the short-channel MOSFET, where intrinsic refers to the capacitances of the active device itself and not the fringing or overlap capacitances, and the analysis of other devices such as the MODFET (Modulation Doped Field Effect Transistor).

8.5a MOSFET Capacitances

The intrinsic capacitances of the MOSFET are calculated from the charges Q_G , Q_M and Q_B , where: Q_G is the total charge stored on the intrinsic part of the gate directly above the channel; Q_M is the total mobile charge in the channel; and Q_B is the total depletion charge in the bulk that is associated with the intrinsic MOSFET. If the effect of majority carriers within the

depletion region is neglected by making the depletion approximation, then charge neutrality implies that

$$Q_G + Q_M + Q_B = 0 \quad (8.5.1)$$

To proceed further the quasi-static approximation must be made which assumes that the charges on the gate, in the channel, and in the bulk, can respond to a much higher frequency than the frequency of the signals applied to the gate, source, drain and bulk terminals. Thus the distribution of charges that determine Q_G , Q_M and Q_B is assumed to be negligibly different from the charge distributions that would exist under steady-state conditions if the same biases were applied. Hence by making the quasi-static approximation Q_G , Q_M and Q_B may be calculated from the two-dimensional potential $\psi(x,y)$ calculated under steady-state conditions, which is the potential that results from the analytical solution presented in this thesis and is given by (4.4.39).

The gate charge Q_G is determined from $\psi(x,y)$ by integrating the normal displacement in the gate oxide. Thus Q_G may be calculated from

$$Q_G = W C_{ox} \int_0^L [V_{GB} - \psi(0,y)] dy \quad (8.5.2)$$

providing that Laplace's equation is assumed to be one-dimensional in the gate oxide, which for thin gate oxides is a good approximation. Since the sheet charge approximation was made in the analytical solution for $\psi(x,y)$, which confines the

mobile charge to an infinitely thin sheet at the silicon surface, the depletion charge per unit area at any point in the channel is equal to $-\epsilon_s \partial\psi/\partial x(0,y)$. So the total depletion charge is found by integrating this quantity over the channel. Hence:

$$Q_B = -W \epsilon_s \int_0^L \frac{\partial\psi}{\partial x}(0,y) dy \quad (8.5.3)$$

Having determined Q_G from (8.5.2) and Q_B from (8.5.3), the total mobile charge Q_M is found by a simple subtraction using (8.5.1). Thus the three integrated charges Q_G , Q_M and Q_B can be determined from the two-dimensional potential given by (4.4.39).

To calculate the intrinsic MOSFET capacitances the total mobile charge Q_M must be divided into a part associated with the source Q_{MS} and a part associated with the drain Q_{MD} . The way this division is made is arbitrary, but it has been found that simply choosing the division point to be the middle of the channel (i.e. $y = L/2$) works well [76]. Upon dividing Q_M there are now four total charges Q_G , Q_{MS} , Q_{MD} and Q_B . Differentiating these charges with respect to the four terminal voltages in either an analytical or numerical fashion yields 16 capacitances, but only 9 are unique. This is because one of the total charges may be chosen as the sum of the others using (8.5.1), and one of the terminals must be used as a reference leaving only three independent voltages. Hence the 9 independent capacitances arise because three independent total charges are differentiated by three independent voltages. These 9 capacitances comprise a complete capacitance model of the intrinsic MOSFET.

Constructing a capacitance model from the general solution of Poisson's equation given by (4.4.39) would mark the first time that a two-dimensional analytical capacitance model of the short-channel MOSFET has been produced. Such a model would be computationally far faster than calculating the capacitances from numerical models. It also would not have the numerical problems caused by using a finite grid that produces considerable numerical scatter in the total charge values, and hence makes it difficult to accurately take the numerical derivatives that are necessary to calculate the capacitances. In addition, an analytical capacitance model may offer more physical insight simply because of its analytical formulation.

8.5b MODFET Modelling

The analytical solutions of Poisson's equation presented in this thesis could also be applied to modelling the behavior of the MODFET. The MODFET is a very fast device constructed from layers of AlGaAs on a GaAs substrate. The structure of the device is shown in Figure 8.5.1 which has been reproduced from [71]. Its basic principle of operation is that electrons are ionized from the donor atoms in the n-type AlGaAs layer and then move to the quantum well at the interface between the undoped AlGaAs and GaAs. The quantum well is caused by the difference in band gaps between AlGaAs and GaAs. The carriers then move in the quantum well as a two-dimensional electron gas under the influence of an applied drain to source voltage. The device is very fast because the electrons in the quantum well do not

Figure 8.5.1. Physical structure of the MODFET reproduced from [71].

experience impurity scattering, and hence move at essentially the same speed as they would in the bulk. The I_D versus V_{DS} characteristics of the MODFET look very similar to those of a MOSFET.

To model the electrical behavior of the MODFET it is necessary to solve for the two-dimensional potential in the vicinity of the two-dimensional electron gas. A model of this type has been derived by Kim and Roblin [71] and shows reasonably good agreement between measured and simulated results. They obtained the two-dimensional potential by solving a two-dimensional Poisson's equation using the method of Laplace reduction and separation of variables just as was done in this thesis. It is possible that applying some of the techniques developed in this thesis could improve the speed and accuracy of Kim and Roblin's model. The detailed investigation of this possibility is a subject for further research.

CHAPTER 9

CONCLUSION

The objective of the research presented in this thesis was to construct a two-dimensional analytical model of the short-channel MOSFET. Because of its analytical formulation, such a model is computationally much faster than the existing numerical models. This speed widens the range of practical applications of two-dimensional short-channel MOSFET models, particularly in the areas of circuit simulation and statistical simulation. A fast analytical model would also be useful for process and device design, and drawing conclusions about the physical behavior of the short-channel MOSFET.

The minimum necessary to create a general model of the short-channel MOSFET, is to simultaneously solve a two-dimensional Poisson's equation and a one-dimensional current continuity equation for the geometry and boundary conditions of this device. Such a solution produces a model that is valid in the subthreshold, triode and saturation regimes. A specialized subthreshold model may be created by solving only a two-dimensional Poisson's equation because the small carrier densities in this regime permit Poisson's equation to be decoupled from the current continuity equation. Once the subthreshold solution of Poisson's equation is found, current continuity may be applied to calculate the drain current. A necessary consideration in solving Poisson's equation and the

current continuity equation, is to include the effect of a field dependent mobility and a nonuniform substrate doping profile, as these are important effects for a short-channel MOSFET.

The two-dimensional analytical solution of Poisson's equation in subthreshold was found in this thesis by reducing Poisson's equation to Laplace's equation using a trial function that corresponds to the ideal long-channel solution, and then solving Laplace's equation by the separation of variables technique. The drain current was then calculated from the resulting two-dimensional potential using a novel interpretation of Gummel's algorithm which is derived from the condition of current continuity. A closed form expression for the short-channel threshold voltage was also obtained from the two-dimensional potential. The solution for the two-dimensional potential has the capability to deal with an arbitrary substrate doping profile, and the expression for the subthreshold drain current includes the effect of a field dependent mobility. The short-channel threshold voltage model was validated by direct comparison with the results of simulations done using the numerical model MINIMOS [33]. Good agreement was found between analytical and numerical threshold voltages. The subthreshold current model was checked by comparing analytically calculated currents to currents measured on carefully characterized devices, and good agreement was obtained.

The simultaneous analytical solution of a two-dimensional Poisson's equation and a one-dimensional current continuity equation, necessary to derive a general model of the

short-channel MOSFET, was obtained using methods that built upon the experience gained with the subthreshold solution. As before, Poisson's equation was solved by reducing it to Laplace's equation using a trial function that is the two-dimensional potential in an ideal long-channel MOSFET. Laplace's equation was then solved using the separation of variables technique. The principle of current continuity was invoked as a defining condition for the Fourier coefficients in the Laplace solution. This was first done using an iterative function technique that was found to be computationally far too slow. It was then found that the Fourier coefficients that provide for current continuity could be calculated by an elegant Fourier analysis technique. This technique reduces the problem to N nonlinear equations in N unknowns that can be solved using Newton and quasi-Newton techniques. The Fourier analysis technique reduces the computation time to between 0.2 and 0.35 CPU seconds which is 5 to 300 times faster than for the one-dimensional current continuity equation level of MINIMOS. The solution of Poisson's equation and the current continuity equation used in the general model assumes a constant mobility and a uniform substrate doping to limit the analytical complexity. The equations and methods necessary to extend the solution so that it includes the effect of a field dependent mobility and a nonuniform substrate doping profile have been included in the thesis but not completed or tested. It was found that the two-dimensional solution of Poisson's equation and the current continuity equation permits the derivation of a closed form equation for the drain current of

a short-channel MOSFET. This expression is the two-dimensional analog of the one-dimensional equation derived by Brews [4]. It was tested by comparing analytically calculated I_D versus V_{DS} and I_D versus V_{GS} characteristics to the results of MINIMOS simulations. The good agreement that was found indicates that the general model of the short-channel MOSFET is accurate:

The advances described in this thesis have opened several new avenues of research. The three most promising areas appear to be: a more general application of the Fourier analysis technique; the calculation of short-channel capacitances from the general analytical solution of Poisson's equation and the current continuity equation; and an application of the techniques developed in this thesis to the simulation of other devices such as MODFET's. Thus it is hoped that not only will the research presented in this thesis prove to be useful, but also that it will stimulate more useful research in these and other areas.

REFERENCES

1. H.K.J. Ihantola and J.L. Moll, "Design Theory of a Surface Field Effect Transistor", Solid-State Electronics, Vol. 7, pp. 423-430, 1964.
2. H.C. Pao and C.T. Sah, "Effects of Diffusion Current on Metal-Oxide (Insulator) - Semiconductor Transistors", Solid-State Electronics, Vol. 9, pp. 927-937, 1966.
3. R.F. Pierret and J.A. Shields, "Simplified Long-Channel MOSFET Theory", Solid-State Electronics, Vol. 26, pp. 143-147, Feb. 1983.
4. J.R. Brews, "A Charge-Sheet Model of the MOSFET", Solid-State Electronics, Vol. 21, pp. 345-355, 1978.
5. J.R. Brews, W. Fichtner, E.H. Nicollian and S.M. Sze, "Generalized Guide for MOSFET Miniaturization", IEEE Electron Device Letters, Vol. EDL-1, pg. 2, 1980.
6. H.W. Loeb, R. Andrew and W. Love, "Applications of 2-Dimensional Solutions of the Shockley-Poisson Equation to Inversion-Layer M.O.S.T. Devices", Electronics Letters, Vol. 4, pp. 352-354, July 1968.
7. G.A. Armstrong, J.A. Magowan and W.D. Ryan, "Two-Dimensional Solution of the D.C. Characteristics for the M.O.S.T.", Electronics Letters, Vol. 5, pp. 406-408, July 1969.
8. F.H. De La Moneda, "Threshold Voltage from Numerical Solution of the Two-Dimensional MOS Transistor", IEEE Trans. on Circuit Theory, Vol. CT-20, pp. 666-673, Nov. 1973.
9. C.L. Wilson and J.L. Blue, "Two-Dimensional Finite Element Charge-Sheet Model of a Short-Channel MOS Transistor", Solid-State Electronics, Vol. 25, pp. 461-477, June 1982.
10. L.D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's", Solid-State Electronics, Vol. 17, pp. 1059-1063, 1974.
11. G.W. Taylor, "The Effects of Two-Dimensional Charge Sharing on the Above-Threshold Characteristics of Short-Channel IGFETs", Solid-State Electronics, Vol. 22, pp. 701-717, 1979.
12. K.N. Ratnakumar and J.D. Meindl, "Short-Channel MOST Threshold Voltage Model", IEEE Journal of Solid-State Circuits, Vol. SC-17, pp. 937-948, Oct. 1982.

13. P.K. Chatterjee and J.E. Leiss, "An Analytic Charge-Sharing Predictor Model for Submicron MOSFETs", International Electron Devices Meeting, Washington D.C., pp. 28-33, 1980.
14. P.P. Guebels and F. Van De Wiele, "A Small Geometry MOSFET Model for CAD Applications", Solid-State Electronics, Vol. 26, pp. 267-273, April 1983.
15. L.M. Dang, "A Simple Current Model for Short-Channel IGFET and Its Application to Circuit Simulation", IEEE Trans. on Electron Devices, Vol. ED-26, pp. 436-445, April, 1979.
16. A.R. Boothroyd, "MOSFET Modelling for CAD in an Industrial Environment", The Third International Conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits, Galway, Ireland, pp. 12-17, June 1983.
17. A.L. Silburt, R.C. Foss and W.F. Petrie, "An Efficient MOS Transistor Model for Computer-Aided Design", IEEE Trans. on Computer-Aided Design, Vol. CAD-3, pp. 104-111, Jan. 1984.
18. W. Maly and A.J. Strojwas, "Statistical Simulation of the IC Manufacturing Process", IEEE Trans. on Computer-Aided Design, Vol. CAD-1, pp. 120-131, July 1982.
19. S.R. Nassif, A.J. Strojwas and S.W. Director, "FABRICS II: A Statistically Based IC Fabrication Process Simulator", IEEE Trans. on Computer Aided Design, Vol. CAD-3, pp. 40-46, Jan. 1984.
20. Y. Aoki, T. Toyabe, S. Asai and T. Hagiwara, "CASTAM: A Process Variation Analysis Simulator for MOS LSI's", IEEE Trans. on Electron Devices, Vol. ED-31, pp. 1462-1467, Oct. 1984.
21. V.G.K. Reddi and C.T. Sah, "Source to Drain Resistance Beyond Pinch-Off in Metal-Oxide-Semiconductor Transistors (MOST)", IEEE Trans. on Electron Devices, Vol. ED-12, pp. 139-141, March 1965.
22. D. Frohman-Bentchkowsky and A.S. Grove, "Conductance of MOS Transistors in Saturation", IEEE Trans. on Electron Devices, Vol. ED-16, pp. 108-113, Jan. 1969.
23. Y.A. El-Mansy and A.R. Boothroyd, "A Simple Two-Dimensional Model for IGFET Operation in the Saturation Region", IEEE Trans. on Electron Devices, Vol. ED-24, pp. 254-262, March 1977.
24. C.G.B. Garrett and W.H. Brattain, "Physical Theory of Semiconductor Surfaces", Physical Review, Vol. 99, pp. 376-387, July 1955.

25. S. Liu, B. Hoefflinger and D.O. Pederson, "Interactive Two-Dimensional Design of Barrier-Controlled MOS Transistors", IEEE Trans. on Electron Devices, Vol. ED-27, pp. 1550-1558, Aug. 1980.
26. J.A. Greenfield and R.W. Dutton, "Nonplanar VLSI Device Analysis Using the Solution of Poisson's Equation", IEEE Trans. on Electron ~~Devices~~, Vol. ED-27, pp. 1520-1532, Aug. 1980.
27. D.B. Scott and S.G. Chamberlain, "A Calibrated Model for the Subthreshold Operation of a Short-Channel MOSFET Including Surface States", IEEE Journal of Solid-State Circuits, Vol. SC-14, pp. 633-644, June 1979.
28. C.R. Viswanathan, B.C. Burkey, G. Lubberts and T.J. Tredwell, "Threshold Voltage in Short-Channel MOS Devices", Vol. ED-32, pp. 932-940, May 1985.
29. A. Husain and S.G. Chamberlain, "Three-Dimensional Simulation of VLSI MOSFET's: The Three-Dimensional Simulation Program WATMOS", Vol. ED-29, pp. 631-638, April 1982.
30. H.K. Gummel, "A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations", Vol. ED-11, pp. 455-465, Oct. 1964.
31. C.K. Kim and E.S. Yang, "An Analysis of Current Saturation Mechanism of Junction Field-Effect Transistors", IEEE Trans. on Electron Devices, Vol. ED-17, pp. 120-127, Feb. 1970.
32. D. Vandorpe, J. Borel, G. Merckel and P. Saintot, "An Accurate Two-Dimensional Numerical Analysis of the MOS Transistor", Solid-State Electronics, Vol. 15, pp. 547-557, 1972.
33. S. Selberherr, A. Schutz and H.W. Potzl, "MINIMOS - A Two-Dimensional MOS Transistor Analyzer", IEEE Trans. on Electron Devices, Vol. ED-27, pp. 1540-1550, Aug. 1980.
34. W. Hansch and S. Selberherr, "MINIMOS 3: A MOSFET Simulator that Includes Energy Balance", IEEE Trans. on Electron Devices, Vol. ED-34, pp. 1074-1078, May 1987.
35. T. Toyabe, K. Yamaguchi, S. Asai and M.S. Mock, "A Numerical Model of Avalanche Breakdown in MOSFET's", IEEE Trans. on Electron Devices, Vol. ED-25, pp. 825-832, July 1978.
36. T. Toyabe and S. Asai, "Analytical Models of Threshold Voltage and Breakdown Voltage of Short-Channel MOSFET's Derived from Two-Dimensional Analysis", IEEE Trans. on Electron Devices, Vol. ED-26, pp. 453-460, April 1979.

37. T.W. Tang, Q.L. Zhang and D.H. Navon, "Analytical Model for Predicting Threshold Voltage in Submicrometer-Channel MOSFET's", IEEE Trans. on Electron Devices, Vol. ED-32, pp. 1890-1893, Sept. 1985.
38. S.K. Mehta and R. Muralidharan, "Analytical Model for Threshold Voltage of Ion-Implanted Short-Channel IGFET's", IEEE Trans. on Electron Devices, Vol. ED-33, pp. 1073-1074, July 1986.
39. D.R. Poole and D.L. Kwong, "Two-Dimensional Analytical Modeling of Threshold Voltages of Short-Channel MOSFET's", IEEE Electron Device Letters, Vol. EDL-5, pp. 443-446, Nov. 1984.
40. D.R. Poole and D.L. Kwong, "Analytical Modeling of the Subthreshold Current in Short-Channel MOSFET's", IEEE Electron Device Letters, EDL-7, pp. 340-343, June 1986.
41. T. Skotnicki, "Quasi-Two-Dimensional Analytical Solution of Poisson Equation in Arbitrarily Doped Short-Channel MOSFET", Electronics Letters, Vol. 19, pp. 797-798, Sept. 1983.
42. T. Skotnicki and W. Marciniak, "A New Approach to Threshold Voltage Modelling of Short-Channel MOSFETs", Solid-State Electronics, Vol. 29, pp. 1115-1127, Nov. 1986.
43. J.D. Kendall and A.R. Boothroyd, "A Two-Dimensional Analytical Threshold Voltage Model for MOSFET's with Arbitrarily Doped Substrates", IEEE Electron Device Letters, Vol. EDL-7, pp. 401-403, July 1986.
44. J.R. Pfiester, J.D. Shott and J.D. Meindl, "Performance Limits of CMOS ULSI", IEEE Trans. on Electron Devices, Vol. ED-32, pp. 333-343, Feb. 1985.
45. T.N. Nguyen, "Small-Geometry MOS Transistors: Physics and Modeling of Surface- and Buried-Channel MOSFETs", Ph.D. Dissertation, Stanford University, Stanford, California, Aug. 1984.
46. V. Marash and R.W. Dutton, "Submicron 2D MOS Modeling", International Conference on Computer Aided Design, Santa Clara, California, pp. 476-479, Nov. 1986.
47. A.R. Boothroyd (private communication).
48. M.R. Pinto, C.R. Rafferty and R.W. Dutton, "PISCES II: Poisson and Continuity Equation Solver", Stanford Electronics Lab, Stanford University, Sept. 1984.
49. B.R. Chawla and H.K. Gummel, "Transition Region Capacitance of Diffused p-n Junctions", IEEE Trans. on Electron Devices, Vol. ED-18, pp. 178-195, March 1971.

50. J.D. Kendall, "A Punchthrough Current Model for Short-Channel MOSFETs", M.Eng. Dissertation, Carleton University, Ottawa, Canada, Sept. 1983.
51. J. Whitfield, "A Modification on "An Improved Method to Determine MOSFET Channel Length"", IEEE Electron Device Letters, Vol. EDL-6, pp. 109-110, March, 1985.
52. J.C. Irvin, "Resistivity of Bulk Silicon and of Diffused Layers in Silicon", Bell System Technical Journal, Vol. 41, pg. 387, 1962.
53. R. Theriault (private communication).
54. D.A. Smith (private communication).
55. J. Kolk (private communication).
56. W. Van Gelder and E.H. Nicollian, "Silicon Impurity Distribution as Revealed by Pulsed MOS C-V Measurements", Journal of the Electrochemical Society, Vol. 118, pp. 138-141, Jan. 1971.
57. J.R. Brews, "Threshold Shifts Due to Nonuniform Doping Profiles in Surface Channel MOSFET's", IEEE Trans. on Electron Devices, Vol. ED-26, pg. 1696, 1979.
58. R.W. Coen and R.S. Muller, "Velocity of Surface Carriers in Inversion Layers on Silicon", Solid-State Electronics, Vol. 23, pp. 35-40, 1980.
59. C.F. Gerald, "Applied Numerical Analysis", Second Edition, Addison-Wesley Publishing Co., Reading, Massachusetts, 1978.
60. A.H. Stroud and D. Secrest, "Gaussian Quadrature Formulas", Prentice-Hall, Englewood Cliffs, New Jersey, 1966.
61. T.N.L. Patterson, "On High Precision Methods for the Evaluation of Fourier Integrals with Finite and Infinite Limits", Numerische Mathematik, Vol. 27, pp. 41-52, 1976.
62. A. Alaylioglu, G.A. Evans and J. Hyslop, "The Use of Chebyshev Series for the Evaluation of Oscillatory Integrals", Computer Journal, Vol. 19, pp. 258-267, 1976.
63. R.K. Littlewood and V. Zakian, "Numerical Evaluation of Fourier Integrals", Journal of the Institute of Mathematics and its Applications, Vol. 18, pp. 331-339, 1976.
64. G. Blanch, "Numerical Evaluation of Continued Fractions", SIAM Review, Vol. 6, pp. 383-421, Oct. 1964.

65. J.M. Ortega and W.C. Rheinboldt, "Iterative Solution of Nonlinear Equations in Several Variables", Academic Press, New York, 1970.
66. E. Zeidler, "Nonlinear Functional Analysis and its Applications I Fixed Point Theorems", Springer-Verlag, Berlin, 1986.
67. C.G. Broyden, "A Class of Methods for Solving Nonlinear Simultaneous Equations", Mathematics of Computation, Vol. 19, pp. 577-593, 1965.
68. J.E. Dennis and J.E. More, "Quasi-Newton Methods, Motivation and Theory", SIAM Review, Vol. 19, pp. 46-89, 1977.
69. G.P. Tolstov, "Fourier Series", translated by R.A. Silverman, Prentice-Hall, Englewood Cliffs, New Jersey, 1962, (Theorem 2, pp. 137-139).
70. I. Gladwell and R. Wait, "A Survey of Numerical Methods for Partial Differential Equations", Clarendon Press, Oxford, pp. 264-274, 1979.
71. Y.M. Kim and P. Roblin, "Two-Dimensional Charge-Control Model for MODFETs", IEEE Trans. on Electron Devices, Vol. ED-33, pp. 1644-1651, Nov. 1986.
72. S.M. Sze, "Physics of Semiconductor Devices", Second Edition, John Wiley and Sons, New York, 1981.
73. A.I. Akinwande, C.P. Ho and J.D. Plummer, "Experimental Determination of the Temperature Dependence of Argon Annealed Fixed Oxide Charge at the Si/SiO₂ Interface", Applied Physics Letters, Vol. 45, pp. 263-265, Aug. 1984.
74. N.G. Tarr (private communication).
75. W.M. Werner, "The Work-Function Difference of the MOS-System with Aluminum Field Plates and Polycrystalline Silicon Field Plates", Solid-State Electronics, Vol. 17, pg. 769, 1974.
76. A.R. Boothroyd (private communication).

END

1|9.0|8.88

FIN