A Surface-Potential-Based Compact Model for Partially Depleted Silicon-On-Insulator MOSFETs

by James Benson

Thesis, April 2009

UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING AND APPLIED SCIENCE
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

A SURFACE-POTENTIAL-BASED COMPACT MODEL FOR PARTIALLY DEPLETED SILICON-ON-INSULATOR MOSFETS

by James Benson

Thesis, April 2009.

To my parents, Leslie and Zora.

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

A SURFACE-POTENTIAL-BASED COMPACT MODEL FOR PARTIALLY DEPLETED SILICON-ON-INSULATOR MOSFETS

by James Benson

With the continuous scaling of CMOS technologies, Silicon-on-Insulator (SOI) technologies have become more competitive compared to bulk, due to their lower parasitic capacitances and leakage currents. The shift towards high frequency, low power circuitry, coupled with the increased maturity of SOI process technologies, have made SOI a genuinely cost-effective solution for leading edge applications.

The original STAG2 model, developed at the University of Southampton, UK, was among the first compact circuit simulation models to specifically model the behaviour of Partially-Depleted (PD) SOI devices. STAG2 was a robust, surface-potential based compact model, employing closed-form equations to minimise simulation times for large circuits. It was able to simulate circuits in DC, small signal, and transient modes, and particular care was taken to ensure that convergence problems were kept to a minimum.

In this thesis, the ongoing development of the STAG model, culminating in the release of a new version, STAG3, is described. STAG3 is intended to make the STAG model applicable to process technologies down to 100nm. To this end, a number of major model improvements were undertaken, including: a new core surface potential model, new vertical and lateral field mobility models, quantum mechanical models, the ability to model non-uniform vertical doping profiles, and other miscellaneous effects relevant to deep sub-micron devices such as polysilicon depletion, velocity overshoot, and the reverse short channel effect.

As with the previous versions of STAG, emphasis has been placed on ensuring that model equations are numerically robust, as well as closed-form wherever possible, in order to minimise convergence problems and circuit simulation times. The STAG3 model has been evaluated with devices manufactured in PD-SOI technologies down to 0.25μm, and was found to give good matching to experimental data across a range of device sizes and biases, whilst requiring only a single set of model parameters.

Acknowledgements

It's a sure sign that a PhD has gone well over time when your first instinct is to break up your acknowledgements into separate sub-sections. In the time since I began this research, I have moved house five times, worked in four different jobs, and seen three generations of PhD students come and go at two universities. Before any more time elapses, let us begin the credits. Profound apologies for any omissions; I'm afraid my memory isn't what it used to be.

I'll start with a big thank you to my supervisor Bill Redman-White. Despite many setbacks, he never lost faith in the project, and he went to great lengths to build collaborations and secure funding in the early years, during what was a turbulent period in the industry. When developing a compact model, it's always useful to see things from the perspective of the designer, and Bill's expertise provided a useful counterpoint to my own physics-based 'side on' view of devices. And while I never got the hang of circuit design, it certainly wasn't due to any lack of encouragement on his part!

I was lucky enough to share an office with two very capable and dedicated PhD students, Nele d'Halleweyn and Ketan Mistry. Nele's similar research interests meant that I always had a sounding board for my ideas, and our conversations provided me with a great deal of insight into the subtleties of compact model development. Ketan provided advice on everything from typesetting LaTeX documents to building measurement gear, and his help was greatly appreciated.

I also had a great deal of help from some of my fellow RAs. Mike Lee, the original author of the STAG model, made sure to pass on as much of his knowledge as he could, and got me off to a good start. Bernard Tenbroek introduced me to device characterisation, and his small-signal measurement setup continued to see use long after his departure. Craig Easson worked with me on the DERA contract until he too was lured to California.

Various partners and collaborators provided us with funding, processed devices, and measurement data. In particular, I'd like to thank Mike Uren of DERA Malvern, Jean-Luc Pelloie and Olivier Faynot of CEA-LETI Grenoble, and Mike Liu of Honeywell Semiconductors. I would also like to express my appreciation to Henri Kemhadjian and Peter Ashburn for keeping me gainfully employed when my research contract came to an end.

Other notables include John Amy and Pinder Sant, the epi gurus, Dave Batt, with whom I had a very productive collaboration writing the software for the EPI4 machine, and Glenys Howe, Lucia Hewett, and Angie Mo, who helped me out more times than I can count!

Socially, I've been fortunate to have made many good friends in each of the places I've lived, and without them I doubt I would have emerged from this whole business with my sanity intact. Since I've been getting around a bit, I'll list them by location!

Southampton: Thanks go to my old partners in crime Mir Mokhtari and Becky Neal, my good friend Elena Koukharenko for all the long chats over coffee, and of course Gabriela Dilliway, who has been a constant presence throughout and who taught me more about this crazy industry than anyone else. A special mention for Rona Barnedo, for all the great times both in and out of the salsa bars. Also, a big shout out to all members of the volleyball team 'The Random Collection', as well as all the other teams who helped make the volleyball league such a success. And of course, Roz Painter, for a wonderful two years, and for being my muse when it really mattered.

Guildford: The Denzil Road posse deserve a special mention for being quite possibly the coolest, craziest, most entertaining bunch of housemates ever gathered under one roof. Rob McCracken, Faye Stacey, Natalie Biddle, Mark Prentice, Shelley Taylor, the ever present Darren Arnold, and many others in all their strangeness. May the Denzil Road parties continue for many years to come!

Not to be outdone, the Surrey University postgrad community was one of the main reasons I stayed so long. Claire Mercier, Melanie Webb, Justin Hamilton, Daphnee Pushparajah, Laura Hyman, Tim Sinnamon, Yvonne Huebner, Liliana Cuenca, Andy Smith and many others all helped to make university life a hugely enjoyable experience for me.

Others that I count myself very fortunate to have met during my time in Surrey include Rebecca Oyabugbe, Mida Aslam (I'm not drunk!), Ashley Browning, Farnaz Behzadmehr, Andreia and Andrea (the famous Brazilian double-act), and Suzi and Harriet for their brief but memorable visit! Last but certainly not least, Fortune Mgbangson, who gave me so much support, encouragement, and inspiration during some difficult times.

Leuven: Some of my colleagues at NXP became good friends, and made my life more fun both in and out of work. Big thanks go to Rob Lander for footie nights and for bringing me over here in the first place, Suzanne Vollenbronck for keeping me out dancing all night, Eero Saarnilehto for asking the big questions (why did you do that?), Benoit Battaillou for being generally evil, and of course Korina Fotopoulou for her huge moral support, for her expert LaTeX assistance, and for putting a roof over my head during the final days of my thesis corrections.

Thanks also go to Tra Pham for all the good times and silly jokes, Lode Devlieghere for keeping me focused on the important things in life, Inge Lontie for showing me the wild side of Leuven, and all the staff of Ron Blacks for keeping the cider flowing. Katrien Segaert, Clara Younan, and Lavinia Butiu all deserve a mention for making my life more interesting, albeit briefly. And finally, thanks to An de Groef (a.k.a. KM), who left me with lots of great memories, and whose charm, good humour, and constant support made my first year in Leuven better than I ever could have expected.

Contents

List of Figures

List of Symbols

a_1	Saturation surface potential cubic coefficient [-]
a_2	Saturation surface potential cubic coefficient [-]
a_3	Saturation surface potential cubic coefficient [-]
a_{vo}	Velocity overshoot proportionality constant [-]
b	Variational parameter [nm^{-1}]
$CGFB0$	Front gate-body overlap capacitance per unit length [F]
$CGFD0$	Front gate-drain overlap capacitance per unit length [F]
$CGFS0$	Front gate-source overlap capacitance per unit length [F]
C_{J0}	Zero bias junction depletion capacitance per unit area [F/cm^2]
C_{jbd}	Body-drain junction depletion capacitance [F]
C_{jbs}	Body-source junction depletion capacitance [F]
C_{of}	Front gate capacitance per unit area [F/cm^2]
C_T	Device thermal capacitance [pJ/K]
C_{T0}	Thermal capacitance associated with first time constant [pJ/K]
C_{T1}	Thermal capacitance associated with second time constant [pJ/K]
C_{T2}	Thermal capacitance associated with third time constant [pJ/K]
C_{T3}	Thermal capacitance associated with fourth time constant [pJ/K]
C_{T4}	Thermal capacitance associated with fifth time constant [pJ/K]
D_{itf}	Fast surface state density [$1/cm^2.eV$]
E_c	Critical lateral electric field for carrier velocity saturation [V/m]
E_{xeff}	Effective vertical electric field [V/m]
E_{yeff}	Effective lateral electric field [V/m]
f_b	Body charge electric field weighting coefficient [-]
f_c	Channel charge electric field weighting coefficient [-]
f_R	Total internal series resistance factor [-]
F	Quotient of drain to source charge densities (q_d/q_s) [-]
F	Channel degeneracy factor [-]
F_{min}	Technology minimum feature size [μm]
FC	Coefficient for forward-bias depletion capacitance model [-]
g_{ds}	Drain-source output conductance [S]
g_m	Front gate transconductance [S]
g_{mb}	Body-source transconductance [S]
G_{cou}	Coulomb scattering mobility contribution [-]
G_{ph}	Phonon scattering mobility contribution [-]
G_{sr}	Surface roughness scattering mobility contribution [-]
G_v	Total scattering mobility contribution [-]
G_{v0}	Total scattering mobility contribution at source end of channel [-]
$\hbar$	Planck's Constant divided by 2π [J.s]
I_{bd}	Body-drain diode current [A]
I_{bs}	Body-source diode current [A]
I_{CH}	Channel current [A]

I_{CHint}	Channel current excluding high field mobility effects [A]
I_{Mdb}	Impact ionisation current in forward mode [A]
I_{Msb}	Impact ionisation current in reverse mode [A]
I_s	Diode diffusion component reverse saturation current [A]
I_{s1}	Diode recombination component reverse saturation current [A]
J_s	Diode diffusion component reverse saturation current density per metre width [A/m]
J_{s1}	Diode recombination component reverse saturation current density per metre width [A/m
J_n	Electron current density per unit area [A/cm^2]
k	Mobility temperature exponent [-]
k_B	Boltzmann constant [J/K]
K_{SP}	Short-hand notation [-], see Equation (2.41)
L	Device channel length [m]
L'	Length of GCA region [m]
l_d	Length of drain region [m]
l_m	Ionisation length [m]
l_x	Short channel channel length modulation parameter [m]
L_{eff}	Effective channel length [m]
L_{m1}	Ionisation length bias coefficient [m/V]
L_{m2}	Ionisation length bias coefficient [m/V^2]
L_{meff}	Effective ionisation length [m]
LD	Source/drain lateral diffusion [m]
L_{DH}	Debye-Huckel length [$\mathring{A}$]
L_s	Effective fixed charge screening length [$\mathring{A}$]
L_{th}	Characteristic carrier thermal length [$\mathring{A}$]
M	Avalanche multiplication factor due to impact ionisation [-]
M	Junction grading coefficient [-]
m_e	Electron effective mass
M_{SP}	Charge linearisation empirical fitting parameter [-]
n	Free electron concentration [cm^{-3}]
N_B	Body doping concentration [cm^{-3}]
N_C	Channel carrier concentration [cm^{-3}]
N_P	Polysilicon gate doping concentration [cm^{-3}]
P	Device power dissipation [W]
PB	Junction built-in potential [V]
q	Electron charge [C]
q_b	Body charge per unit area [C/cm^2]
q_{b0}	Body charge density at source end of channel [C/cm^2]
Q_{B1}	Body charge in GCA region [C]
Q_{B2}	Body charge in saturated drain region [C]
q_c	Channel charge per unit area [C/cm^2]
q_{c0}	Channel charge density at source end of channel [C/cm^2]
Q_{CH1}	Channel charge in GCA region [C]
Q_{CH2}	Channel charge in saturated drain region [C]
q_d	Channel charge density at drain end [C/cm^2]
Q_{D1}	Drain charge in GCA region [C]
Q_G	Front gate charge [C]
q_s	Channel charge density at source end [C/cm^2]
q_{s0}	Channel screening charge density [C/cm^2]
Q_{S1}	Source charge in GCA region [C]
q_{tot}	Total charge density in silicon film [C/cm^2]
R_D	Drain series resistance [Ω]

R_{DW}	Width scaling drain series resistance [$\Omega.\mu$m]
R_S	Source series resistance [Ω]
R_{SW}	Width scaling source series resistance [$\Omega.\mu$m]
R_P	Projected range of doping implant [nm]
R_T	Device thermal resistance [K/W]
R_{T0}	Thermal resistance associated with first time constant [K/W]
R_{T1}	Thermal resistance associated with second time constant [K/W]
R_{T2}	Thermal resistance associated with third time constant [K/W]
R_{T3}	Thermal resistance associated with fourth time constant [K/W]
R_{T4}	Thermal resistance associated with fifth time constant [K/W]
S	Saturation factor accounting for mobility effects [-]
T	Device temperature (including self-heating effects) [K]
T_{amb}	Ambient temperature [K]
t_b	Silicon film thickness [m]
t_{box}	Back gate oxide thickness [m]
t_{ox}	Front gate oxide thickness [m]
t_{oxeff}	Effective front gate oxide thickness (accounting for quantum effect) [m]
V_{BS}	Body-source voltage [V]
V_{cb}	Channel-body potential [V]
V_{Dex}	Drain-source voltage used to extract V_{Tex} [V]
V_{DB}	Drain-body voltage [V]
V_{DS}	Drain-source voltage [V]
V_{Dsat}	Drain saturation voltage [V]
V_{FB}	Flat-band voltage [V]
V_{FB}^f	Front flat-band voltage [V]
V_g	Short-hand notation [V], see Equation (2.35)
V_{gy}	Numerically limited form of V_g [V]
V_{GBT}	Short-hand notation [V], see Equation (2.31)
V_{GfB}	Front gate-body voltage
V_{GfS}	Front gate-source voltage
V_{GT}	Short-hand notation [V], see Equation (2.28)
V_P	Short channel channel length modulation parameter [V]
v_{sat}	Carrier saturation velocity [cm/s]
v_{sat_vo}	Carrier saturation velocity (accounting for velocity overshoot) [cm/s]
V_{SB}	Source-body voltage [V]
V_{Tex}	Extracted threshold voltage [V]
V_{TH}	Threshold voltage [V]
W	Channel width [m]
x_d	Depletion depth [nm]
W_{eff}	Effective channel width [m]
y_{out}	Drain admittance [S]
y_{sat}	Short-hand notation [-], see Equation (4.21)
Z_{bd}	Body-drain network impedance [Ω]
Z_{bs}	Body-source network impedance [Ω]
α	Short-hand notation [-], see Equation (2.28)
α_0	Impact ionisation coefficient [cm^{-1}]
α_{cou}	Coulomb scattering coefficient [m^3]
α_{ph}	Phonon scattering coefficient [$(\mathrm{m/V})^{1/3}$]
α_{sr}	Surface roughness scattering coefficient [$(\mathrm{m/V})^2$]
α_θ	Mobility scattering coefficient [cm/V]
β_0	Impact ionisation coefficient [V/cm]

χ	Temperature coefficient of threshold voltage [V/K]
χ_β	Temperature coefficient of impact ionisation parameter [V/(cm.K)]
χ_{FB}	Temperature coefficient of flat-band voltage [V/K]
δ_0	Empirical threshold extraction parameter [-]
δ_s	Body charge linearisation coefficient [$V^{-1/2}$]
δ_p	Body charge linearisation coefficient [$V^{-1/2}$]
ΔE_g	Electron ground state offset from bottom of conduction band [eV]
ΔL	Short channel effect parameter [m]
ΔR_P	Standard deviation of R_P [nm]
ΔT	Device temperature rise [K]
ΔW	Narrow width effect parameter [m]
Δz_{CL}	Classical inversion charge centroid offset [m]
Δz_I	Inversion charge centroid offset [m]
Δz_{QM}	Quantum mechanical inversion charge centroid offset [m]
η	Impact ionisation empirical field adjustment parameter [-]
η_d	Diode ideality factor for diffusion component [-]
η_{d1}	Diode ideality factor for recombination component [-]
η_s	Fast surface state coefficient [-]
γ_{eff}	Effective body factor [$V^{1/2}$]
γ_s	Silicon body factor [$V^{1/2}$]
γ_p	Polysilicon gate body factor [$V^{1/2}$]
κ	Thermal conductivity [W/m.K]
λ	Channel length modulation parameter [m/V]
λ_{vo}	Velocity overshoot parameter [V/K]
μ	Carrier mobility [cm^2/V.s]
μ_0	Low field carrier mobility [cm^2/V.s]
μ_n	Electron mobility [cm^2/V.s]
μ_s	Surface carrier mobility [cm^2/V.s]
μ_{xeff}	Effective mobility accounting for vertical field degradation [cm^2/V.s]
μ_{eff}	Effective mobility accounting for carrier velocity saturation [cm^2/V.s]
μ_{cou}	Coulomb scattering limited mobility [cm^2/V.s]
μ_{ph}	Phonon scattering limited mobility [cm^2/V.s]
μ_{sr}	Surface roughness scattering limited mobility [cm^2/V.s]
μ^*	Unscreened mobility per scattering centre per unit area [$(V.s)^{-1}$]
ω	Short-hand notation [V], see Equation (2.53)
ω_{sinv}	Short-hand notation [V], see Equation (2.49)
Φ_F	Fermi potential [V]
Φ_T	Thermal voltage (diode equations) [V]
Φ_t	Thermal voltage [V]
Ψ	Short-hand notation [V], see Equation (4.1)
ψ_s	Surface potential [V]
ψ_{s0}	Surface potential at source end of channel [V]
ψ_{sinv}	Strong inversion surface potential approximation [V]
ψ_{sLsat}	Saturation surface potential [V]
ψ_{sL}	Surface potential at drain end of channel [V]
ψ_{ss}	Surface potential in sub-threshold [V]
ψ_{st}	Single piece surface potential approximation (no high field effects) [V]
ψ_{st0}	Surface potential at source end of channel (no high field effects) [V]
ψ_{stL}	Surface potential at drain end of channel (no high field effects) [V]
ψ_{stqm0}	Surface potential at source end of channel accounting for quantum effects [V]
ψ_{stqmL}	Surface potential at drain end of channel accounting for quantum effects [V]

σ	DIBL parameter [m]
θ	Vertical field mobility degradation [V^{-1}]
ϵ_{ox}	Permittivity of silicon dioxide [$11.7 \times 8.854 \times 10^{-12}$ F/m]
ϵ_{si}	Permittivity of silicon [$3.9 \times 8.854 \times 10^{-12}$ F/m]

List of Acronyms

BJT	Bipolar Junction Transistor
CMOS	Complementary Metal–Oxide–Semiconductor (technology)
DIBL	Drain Induced Barrier Lowering
DOS	Density of States
FD	Fully Depleted
GIDL	Gate Induced Drain Leakage
IC	Integrated Circuit
LDD	Lightly Doped Drain
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
PD	Partially Depleted
RSCE	Reverse Short Channel Effect
SCE	Short Channel Effect
SOI	Silicon-On-Insulator
SOS	Silicon-On-Sapphire
STAG	Southampton Thermal AnaloGue (SOI MOSFET SPICE model)
VLSI	Very Large Scale Integration

Chapter 1

Introduction

1.1 A Review of MOS Technology

1.1.1 Bulk Technology

Despite the success of bulk CMOS in the microelectronics industry, devices manufactured in the conventional way do exhibit some non-ideal behaviour. There have traditionally been two problems in particular which have affected bulk CMOS devices, both of which result from a lack of electrical isolation between the active device and the substrate.

The first undesirable interaction involves a phenomenon known as latch-up [1]. This is unique to CMOS, and occurs through the activation of a PNPN thyristor structure, formed by three PN junctions in close proximity. When this happens, parasitic bipolar action creates a short circuit path between the power terminals, destroying the device. Various measures can be employed to avoid this, including using an oxide isolation wall to break the PNPN parasitic structure, or utilising the twin-tub configuration [1]. Of course, these measures introduce additional process steps, and so raise the cost of producing a wafer. Since it is the parasitic biolar action which governs latch-up, it would follow that sub-micron processes are again more susceptible to this effect, featuring as they do larger parasitic BJT gains. Admittedly, latch-up is no longer the major concern it used to be, since modern rail voltages are often too low to turn on the parasitic PNPN device.

The second problem is the parasitic capacitance between the source/drain regions and the substrate. The effect of this is to slow down device operation, as these capacitors charge and discharge. Increasing the amount of substrate doping results in larger parasitic capacitances; unfortunately, this doping increase is necessary for scaling in modern deep sub-micron processes, in order to prevent punch-through and to help regulate the threshold voltage [2]. As a result, we expect the parasitic capacitance per unit area to increase with decreasing feature size.

Not only does this second parasitic effect become more problematic as MOS devices are scaled down, but new circuit applications now exist which tend to emphasise these prob-

lems. High frequency circuitry, critical to the wireless telecommunication market, requires fast device operation, and parasitic capacitances are extremely undesirable. Thus, any MOS technology which solves these problems, and which provides a viable alternative to bulk CMOS, is worthy of serious consideration. SOI is one such technology.

1.1.2 SOI Technology

SOI (Silicon-on-Insulator) technology has been in existence in some form for almost as long as bulk CMOS. The basic principle behind SOI is as follows: since it is the interaction between the active device and the substrate which causes many undesirable parasitic effects in bulk technology, simply replace the silicon substrate with an insulating substrate, thus electrically decoupling the active device. Initially, the substrate used was made of sapphire, and the technology was called SOS (Silicon-on-Sapphire). Later, it became possible to create good quality interfaces using silicon oxide, which remains the material of choice today. While it has proved very useful in certain niche markets, SOI has never come close to rivalling bulk CMOS in terms of volume of production or breadth of application.

SOI brings a number of benefits to circuit designers. The electrical isolation of the active device area eliminates any possibility of latch-up, whilst also greatly reducing parasitic drain and source capacitances. In addition, only a thin film of silicon is being used for the active device, so the source and drain diffusions will reach to the back oxide, and the p-n junction areas will be smaller. This in turn leads to smaller leakage currents, which makes it particularly useful for high temperature applications. Another benefit of SOI is that it uses a smaller active volume of silicon compared to bulk, making it less susceptible to single event upset caused by high levels of radiation. These two properties account for the usefulness of SOI technology in high temperature and rad hard applications, which for a long time were its main commercial uses.

With the advent of wireless communications, the demand for low power, high frequency systems has increased greatly. Such applications are also good candidates for development in SOI, since they can benefit from the reduced parasitic capacitances in order to achieve higher speed. Furthermore, the combination of reduced junction leakage current and parasitic capacitances mean that SOI CMOS circuits can achieve lower levels of power consumption, both when static and during switching, making SOI an attractive prospect for low power circuitry.

The presence of an electrically isolated active area can cause problems however. The well-documented kink effect [3, 4] occurs when part of the body is left floating, due to the lack of a good electrical contact to the silicon film at the back oxide interface. The resulting unpredicability of the body node voltage can result in complex behaviour in DC, small signal and transient regimes [3, 5, 6]. Additional, less well known small signal phenomena can also be observed [6], which are caused by capacitive coupling of the floating body

node. These floating body effects make SOI behaviour less predictable than bulk, and conventional bulk models cannot predict these behaviours.

Devices in which the body region is not totally depleted when the MOS channel is inverted, and which are thus susceptible to the kink effect, are referred to as *Partially-Depleted SOI* (PD-SOI) devices. Another class of SOI MOS transistor exists, one in which the silicon film is sufficiently thin to give total depletion of the body even at low gate biases. These are called *Fully-Depleted SOI* (FD-SOI) devices. An ideal fully-depleted device is one in which there exists no undepleted quasi-neutral region in the body of the device, electrically isolated from the source and drain junctions. Any carriers injected into the body region below the channel, for instance as a result of impact ionisation, will not be able to remain in the depleted body region, and thus they cannot contribute to forward biasing of the source/body junction. Instead, they will be swept towards the source terminal by the built-in potential across the body-source junction [8]. Furthermore, lowering of the body-source junction barrier allows oppositely charged carriers to diffuse from the source into the body, with the result that recombination will occur before a substantial increase in body potential can occur.

Fully-depleted devices offer the benefits of SOI performance but without the kink effect, and with transconductance behaviour that is superior to both bulk and PD-SOI [7]. A decade ago, mature FD-SOI technologies had not yet been realised, due to limitations in processing techniques. In order to produce a device for which the gate-induced depletion region reaches all the way to the back gate, it is necessary to either reduce the doping level of the silicon film, or else reduce the thickness of the silicon film. The first option is prohibited in short-channel devices, since high doping levels are necessary to suppress short-channel effects. Therefore, it is necessary instead to use thinner silicon films. However, it is more difficult to ensure consistent film thicknesses in production [9]. Nowadays, FD-SOI is the more common technology, with PD-SOI finding use in certain high-performance niche applications.

Finally, SOI technology exhibits self-heating, due to the poor thermal conductivity of the buried oxide layer [10, 11]. Self-heating appears in both types of SOI device, and can degrade device performance. Because the heating is simply due to the power being dissipated from the channel current, the degradation becomes most noticeable at high gate and drain voltages. As with the electrical floating-body effects, this needs to be properly modelled, to avoid unexpected circuit behaviour.

1.1.3 SOI versus bulk

For several decades, SOI has occupied only certain niche markets in the microelectronics industry. While its properties have made it ideal for rad hard and high temperature applications, the difficulty of modelling and compensating for the various floating body effects,

combined with the additional cost compared with the more mature bulk technologies, have meant that SOI has been unable to compete in the core digital and mixed signal VLSI markets.

This situation began to change in the late 1990s, following IBM's announcement of its first commercial microprocessor designed in SOI technology, and a number of examples of commercial digital SOI designs can now be found, i.e. [12]. With the move to deep-submicron technologies, and the growing demand for low-power, high frequency circuit designs, the advantages of SOI over bulk have started to become sufficiently attractive for companies to invest time and money trying to overcome the process and design problems which have previously stifled the exploitation of this technology. The promise of improved performance, either through higher speeds or lower power consumption, has led companies such as Motorola, Texas Instruments and AMD to also start developing commercial SOI-based products.

At present, there is no clear consensus as to whether SOI can provide a long-term alternative to bulk. While it is generally agreed that SOI performance exceeds that of bulk by 10-20% [13], there is much debate as to whether this advantage can be maintained with further technology scaling. Much of the improvement afforded by SOI is a result of the reduction of the junction capacitance. However, as devices are scaled down, the issue of junction capacitance becomes secondary to parasitic capacitances in the gate and interconnects [13]. Furthermore, because floating-body devices are susceptible to the history effect [14], circuit designers must factor in a safety margin when designing circuits in SOI. This means that they cannot utilise the full performance gain of the technology, so that the theoretical improvement of around 15% is reduced to about 10% in practise. As a result, some companies, notably Intel, have been reluctant to commit resources to development of a SOI process.

Regardless of the performance issues, SOI circuit production is still somewhat limited by logistical issues arising from its lack of maturity relative to bulk. SOI wafers still cost more than bulk, and until recently, were not available in the quantities needed for volume production. However, the fact that so many semiconductor maufacturers are now taking SOI seriously means that this situation is likely to improve. It is difficult to see SOI totally replacing bulk over any timescale; the only scenario that could bring this about would be if bulk MOSFETs encountered some fundamental limit in scaling, one to which SOI was immune. Good progress is still being made in the scaling of transistors however, with commercial design work now being done with device gate lengths of 45nm for leading edge products. It is thought that ultimately, SOI might postpone the onset of any performance limit by about one generation, on account of the advantages that it provides over bulk. Of course, as already discussed, this assumes that advances in other areas such as interconnect technology allow it to retain its performance advantages. A plausible prediction would be that SOI will continue to mature and co-exist with bulk processes, with some

companies developing bulk and SOI processes in parallel, and using each for different types of application.

1.2 A Review of MOS Compact Models

We shall now look at the current state of MOS compact modelling. We shall be taking a broader look at available models, encompassing PD-SOI, FD-SOI, and bulk models. Having categorised the various basic model types, we shall then explain how the work described in this thesis fits into this wider picture.

1.2.1 Piece-Wise Models

Earlier generations of compact models are based on the Meyer model [15]. The Meyer model represented the first serious attempt to describe MOSFET characteristics in a relatively simple way, as a function of the applied terminal voltages. The model equations were derived by assuming that the channel current was composed of only a drift component, with the diffusion contribution being ignored. The original Meyer model failed to fulfill the important condition of charge conservation, a shortcoming that was later corrected by the capacitance model developed by Ward and Dutton [16].

Two main difficulties have emerged when using the Meyer formulation as the basis of a compact model. The first problem is that the model equations possess discontinuities. Since we are neglecting the diffusion current, we find a discontinuity when making the transition between strong inversion (where drift does indeed dominate over diffusion) and sub-threshold (where the reverse is true). The same is true for the transition between the strong inversion triode region and saturation. Separate equations need to be introduced to describe sub-threshold and saturation regions, and smoothed numerically. Models which describe different operating regions with different equations are called 'piece-wise' or 'regional' models. It is in fact perfectly possible to eliminate such discontinuities through the application of appropriate smoothing functions, but the need for such measures was less apparent at the time these models were being developed, and so the problem of discontinuities is one that has become associated with piece-wise models.

The second problem is that it is difficult to incorporate new physical effects that result from device scaling. This stems from the empirical nature of these models, and is usually reflected in the model possessing large numbers of empirical fitting parameters. It is not unusual for such models to employ 'binning', whereby each key parameter is given additional dependencies for device length, width and area, thus making model parameter sets large and unwieldy. Fitting for such models is often very good under ideal circumstances, since the extra fitting parameters offer so many degrees of freedom. However, actually optimising such a large number of parameters is difficult and time-consuming, and can

lead to non-physical parameter values being assigned. This in turn greatly reduces the model's predictive power, since there is less likelihood of accuracy outside the optimisation range.

Although it is now standard practise for models to be based on the newer surface potential based approach, there exist some well-known older piece-wise models. Among them is the BSIM3 model, which for a long time was the industry standard. There are versions of BSIM3 for both bulk MOS [17] and SOI [18, 19]. Another well known piece-wise model is the Philips MOS9 model [20]; like BSIM3, these is well established, and can be fitted to give good matching to experimental data.

1.2.2 Surface Potential Models

This newer class of model is effectively considered the de facto standard in compact circuit model development, even if the microelectronics industry has been somewhat slow in the past to actually adopt these models into their design flow [21]. At the core of the surface potential approach is the drift-diffusion approximation, originally developed by Pao and Sah [22]. As with the piece-wise models, the intention is to describe the characteristics of a MOSFET as a function of its applied terminal voltages. However, the inclusion of both drift and diffusion current components leads to a more unified physical description which includes all bias conditions within a single expression. To obtain this expression requires the combined solution of three key equations: the Poisson equation, which relates the electric field to the distribution of charges, the current density equation, which describes the current as being composed of a drift and a diffusion component, and the continuity equation, which maintains the overall carrier flux at zero.

By solving the above equations, an implicit expression for the device surface potential is obtained. The drain current is then expressed as a function of the surface potential at the drain and source ends of the channel. The use of a single consistent implicit equation removes both of the earlier problems at a stroke. The transition between different operating regions is handled automatically, with no discontinuities, and with physical consistency being retained. The situation becomes somewhat less ideal as devices continue to be scaled down. Since the solution to Poisson's equation is obtained in 1-D, the model will lose accuracy as the increasing influence of the lateral electric field creates a 2-D field distribution. Models based on 2-D solutions have been proposed [23] , but most implemented models rely on empirical modifications to compensate for any shortcomings in the 1-D approach, to prevent the equations from becoming too complicated. It should be pointed out that surface potential models still handle this problem much more readily than piece-wise models, and it is now widely accepted that these new models offer superior accuracy, stability, and predictive power.

The expression relating surface potential to the MOSFET terminal voltages is implict -

in other words, no exact closed form solution can be obtained. This leaves the compact model developer with a major decision to make when deciding how to formulate a new model. There are two basic approaches: either solve the exact equation by means of a iterative procedure, or else find an approximate version of the exact expression which does allow for a closed-form solution. We shall call the former group *iterative* models, and the latter group *analytical* models. As far as comparing the two types of model goes, it basically comes down to a trade-off between speed and accuracy.

The iterative models are very precise, with the main benefit coming in the modelling of the weak and moderate inversion regions. This is particularly evident in the transconductance characteristics, since this quantity is a strongly varying function of gate voltage and peaks prior to achieving strong inversion. However, such accuracy can be deceptive. The implicit surface potential equation is a low-field expression, and thus it cannot reproduce real device characteristics by itself. For that, it is necessary to introduce additional expressions to account for the effects of high vertical field mobility degradation, as well as lateral field carrier velocity saturation. Unless these are also modelled to a very high degree of accuracy, some of the benefit of using the exact surface potential equation will be lost.

The most obvious drawback with using an iterative approach is of course speed. Quite how much of a difference this makes is unclear; the past popularity of the BSIM3 models means that until quite recently, most speed benchmarking and comparisons of numerical stability are made against this standard [24, 25]. As has already been pointed out, BSIM3 is a dated model, and most surface potential compact models can expect to outperform it. BSIM3 has therefore been an easy target in the past (but also more readily available for comparison), and not one likely to provide an objective measure of performance by modern standards.

Some information is available on the speed performance of iterative models however, much of it from the group at Hiroshima University. Their model, HiSIM, is one of the most highly developed examples of an iterative model [26], and they have made some effort to benchmark its performance, although some reservations must be expressed about their conclusions. One useful figure of merit in these types of models is the number of iterations needed for the surface potential to converge. For HiSIM, two iterations are typically needed in strong inversion, and three in sub-threshold [25]. This is actually quite a good result, and has been achieved by using predictive algorithms to obtain good initial estimates prior to iteration. However, it still means that the time taken to calculate the source and drain surface potentials is well in excess of that required for closed-form models. Also, it should be understood that when trying to simulate large circuits, the presence of iterative procedures within the model itself can lead to additional convergence problems. Some work has been done in looking at the performance of these models in circuit simulation. The conclusion in [25] was that a well formulated iterative model can actually outperform

an analytical piece-wise model. However, this can probably be attributed to the poor convergence often associated with piece-wise models, and tells us little about the model's performance relative to analytical surface potential models.

Besides HiSIM, other models which fall into this category include MISNAN [27] and the Motorola model [28].

Probably the most widely-used analytic compact model today is the PSP model [29, 30]. The model is the result of a joint collaboration between two well-established compact modelling groups at Philips Semiconductors (now NXP Semiconductors) and Pennsylvania State University. Prior to its inception, each group had developed its own surface potential-based model: MOS11 for Philips, and SP for Pennsylvania State. PSP combines elements from both models, and has now emerged as a serious alternative to BSIM3 and its surface-potential based successor, BSIM4. Prior to PSP, BSIM3 enjoyed such a high level of support from industrial CAD engineers that other models stuggled to challenge it [21].

Other analytic models which include SUSOS [31], STAG2 [32], the Eindhoven model [33], and later versions of SOISPICE [34].

1.3 The STAG Model

The work presented in this thesis relates to the continuing development of the STAG circuit simulation model. STAG was originally conceived and developed at the University of Southampton, in collaboration with the Defence Research and Evaluation Agency (DERA), to allow in-house circuit design work to be carried out. STAG is a compact model for partially depleted SOI MOSFETs, originally implemented in version 3f5 of the Berkeley SPICE circuit simulator. The STAG model has been designed specifically for the reliable simulation of analogue SOI CMOS circuits and includes a rigorous physical treatment of self-heating and floating body effects. Great emphasis has been placed on numerical robustness to prevent convergence problems during the simulation of high-gain analogue circuits. The resulting model uses a surface potential based single-piece formulation with continuous derivatives in all regions of device operation. All model equations are closed form, the rationale being that the simulation time of large circuits would suffer if iterative methods were used. The STAG model is charge conserving and includes static, transient, small-signal, and basic noise models.

The first publicly available version of the model was STAG2, which was released in 1997. While a number of minor iterations were subsequently released, they all shared a common set of core model equations; therefore they will be referred to collectively under the blanket label STAG2. STAG2 was found to provide accurate circuit simulation capabilities for process technologies down to about $0.7\mu m$. STAG2 could be used for DC, small signal, and

transient simulations. A thermal sub-circuit, and associated thermal node, was provided, allowing complex modelling of self-heating effects with single or multiple thermal time constants, as well as the simulation of thermal coupling between devices. The inclusion of impact ionisation models and body-source and body-drain diode sub-circuits enabled simulation of floating body electrical effects associated with PD-SOI devices, such as the kink effect.

Although STAG2 has been shown to be a good, reliable circuit design tool [32,35,36], some of its underlying DC equations are strictly only applicable to long channel devices. When matching simulations to the characteristics of 0.35 and 0.25 μm technologies, it became apparent that there was a limit to the degree of accuracy that could be achieved with the model, even after parameter optimisation. This is especially true when matching was needed over a range of device sizes and bias conditions. A number of major improvements were identified, including new vertical and lateral field mobility models, and the addition of models to include: quantum mechanical effects, non-uniform vertical doping profiles, the reverse short channel effect, and polysilicon depletion.

The model presented in this thesis, STAG3, is the new version of STAG developed for accurate simulation of deep-submicron PD-SOI devices. The principal intention has been to maintain the analytic nature of the model, whilst at the same time improving its accuracy. As before, much effort has gone into ensuring that the model has a high degree of numerical stability.

Before continuing, it is worth noting that much of the new work described in this thesis is not particularly specific to SOI, despite the fact that an SOI model is being used. Rather, the intention is to extend the validity of the STAG PD-SOI model down into the deep sub-micron regime. Most of the changes are just as applicable to bulk MOS as to SOI. The most notable exception to this is the analysis of small-signal capacitance modelling for floating body SOI devices (described in Chapter 8).

1.4 Structure of this thesis

The thesis will begin by looking at the new core low field DC model in STAG3, presented in detail in Chapter 2. A decision has been taken to allow for the possible presence of a non-degenerately doped polysilicon gate, and to include its effect directly within the core surface potential and charge equations. In this way, the effects of any polysilicon depletion on device surface potential are automatically included. This is in contrast to other compact models, which include polysilicon depletion by treating it as a perturbation to the standard surface potential equations. A complete set of new equations are derived, and where appropriate, compared to the more standard treatment. It is shown that the standard surface potential equation can be considered to be a special case of the new expression, corresponding to an degenerately doped gate.

Chapter 3 is the first in a trio of chapters which examines how the presence of high electric fields results in deviations from the ideal low field model, necessitating modification of the surface potential. We begin by detailing the new effective surface mobility model that has been developed for STAG3. This chapter first takes an in-depth look at the state of play of mobility modelling in circuit simulators, as well as covering the physical scattering mechanisms that make an important contribution in the MOS inversion layer. We then explain how expressions have been obtained to express the contribution of each mechanism in terms of the transverse electric field, which in turn allows us to formulate the effective surface mobility in terms of the surface potential.

Chapter 4 explores the additional complications which arise when the effect of the lateral electric field is included. In particular, it is shown that the additional model complexity introduced in Chapter 3 can lead to serious problems when calculating the effects of carrier velocity saturation. We then provide a detailed mathematical formulation which solves this problem. Models for internal series resistance and velocity overshoot are also included in the treatment.

In Chapter 5 we discuss the changes in device behaviour which result from quantisation of the carriers in the inversion layer, due to the presence of high transverse fields. Some simple model equations are proposed to account for these changes.

Chapter 6 examines the section of the model relating to the threshold voltage. We look at the effect of non-uniform body doping, normal and reverse short channel effects, and drain-induced barrier lowering, and show how these effects have been accounted for in the STAG3 model. We present a novel approach to modelling the influence of non-uniform body doping profiles, which provides accuracy without resorting to numerical iteration. Finally, we describe a method for extracting the threshold voltage, by bridging the gap between the well-defined flat band voltage concept used in surface potential models, and the more vaguely defined (but nevertheless very common) threshold voltage extraction procedures.

Chapter 7 completes our treatment of the DC model by looking at the auxilliary model. Improvements made in this area mainly relate to the automatic calculation and scaling of various instance parameters (parameters specific to a particular device). Such changes allow circuit designers to run simulations using default values for thermal parameters, series resistance etc without having to hand calculate values for each variation in device dimensions.

Chapter 8 deals with the charge model. This portion of the model has not needed extensive modifications, since the basic assumptions made for STAG2 still hold for deep sub-micron devices. Where improvements are observed over this earlier version, often this

is due to DC model improvements 'filtering through' to the AC or transient domain. The first parts of this chapter are therefore a review of STAG2, rather than a discussion of new work. The exception to this is the body charge model, for which a detailed treatment is presented. After developing the necessary mathematical formalism, we review the effect of a capacitively coupled floating body node on the small-signal frequency response of SOI devices. We shall see that adopting a conventional bulk MOS approach when modelling body charge capacitances can lead to non-physical characteristics in SOI simulations. Finally, an empirical adjustment to the body charge model is introduced which can be used to compensate for the problem.

In Chapter 9 we evaluate the performance of STAG3 by comparing its output with measured results from a quarter micron technology. While some specific evaluation results are presented in Chapters 2 - 8, here we consider the model in its entirety.

Chapter 10 summarises the conclusions in this thesis, and outlines further improvements that could be made to the STAG model.

References

[1] E S Yang, *Microelectronic Devices*, McGraw-Hill International Editions, 1988.

[2] Y P Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill Book Company, 1987.

[3] J Tihanyi and H. Schlotterer, "Properties of EFSI MOS transistors due to the floating substrate and the finite volume", *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 1017–1023, 1975.

[4] S S Eaton and B Lavevic, "The effect of a floating substrate on the operation of silicon-on-sapphire transistors", *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 907–912, Aug. 1978.

[5] A Wei, M J Sherony, and D A Antoniadis, "Transient behavior of the kink effect in partially-depleted SOI MOSFET's", *IEEE Electron Device Letters*, vol. 16, no. 11, pp. 494–496, Nov. 1995.

[6] R Howes and W Redman-White, "A small-signal model for the frequency-dependent drain admittance in floating-substrate MOSFETs", *IEEE Journal of Solid-State Circuits*, vol. 27, no. 8, pp. 1186–1193, Aug. 1992.

[7] J-P Colinge, "Fully Depleted SOI CMOS for Analog Applications", *IEEE Transactions on Electron Devices*, vol. 45, no. 5, pp. 1010–1016, May. 1998.

[8] J-P Colinge, *Silicon-On-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers, 1991.

[9] G. Shahidi, A. Ajmera, F. Assaderaghi, E. Leobandung, D. Sankus, W. Rausch, D. Schepis, L. Wagner, K. Wu, and B. Davari, "Non-fully-depleted SOI technology for digital logic", in *Proceedings ISSC Conference*, 1999.

[10] L J McDaid, S Hall, P H Mellor, W Eccleston, and J C Alderman, "Physical origin of negative differential resistance in SOI transistors", *Electronics Letters*, vol. 25, no. 13, pp. 827–828, June 1989.

[11] L. T. Su, J. E. Chung, D. A. Antonadis, K. E. Goodson, and M. I. Flik, "Measurement and modeling of self-heating in SOI NMOSFET's transistors", *IEEE Transactions on Electron Devices*, vol. 41, no. 1, pp. 69–74, Jan. 1994.

[12] D H Allen, A G Aipperspach, D T Cox, N V Phan, and S N Storino, "A 0.20μm 1.8V SOI 550MHz 64b PowerPC microprocessor with copper interconnects", in *Proceedings ISSC Conference*, 1999, pp. 438–439.

[13] D Lammers, "CPU Vendors Split Over SOI", *Electronics Times*, pp. 36–37, 24th July 2000.

[14] K. A. Jenkins, J. Y.-C. Sun, and J. Gautier, "History Dependence of Output Characteristics of Silicon-on-Insulator SOI MOSFET's", *IEEE Electron Device Letters*, vol. 17, no. 1, pp. 7–9, 1996.

[15] J.E. Meyer, "MOS Models and Circuit Simulation", *RCA Review*, vol. 32, pp. 42–62, 1971.

[16] D E Ward and R W Dutton, "A Charge-Oriented Model for MOS Transistor Capacitances", *IEEE Journal of Solid-State Circuits*, vol. SC-13, pp. 703–708, 1978.

[17] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CA94720, *BSIM4.2.0 MOSFET Model - User's Manual*, 2001.

[18] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CA94720, *BSIMPD2.2 MOSFET Model User's Manual*, 2001.

[19] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CA94720, *BSIMSOI v2.1 MOSFET Model - Users's Manual for BSIMFD2.1*, 1999.

[20] H C de Graaf and F M Klaassen, *Compact Transistor Modelling for Circuit Design*, Computational Microelectronics. Springer-Verlag, Wien, New York, 1990.

[21] J. G. Fossum, "A Model Too Hot to Handle?", *IEEE Circuit and Devices Magazine*, vol. 18, no. 03, pp. 26–31, May. 2002.

[22] H. C. Pao, and C. T. Sah, "Effects Of Diffusion Current On Characteristics Of Metal-Oxide (Insulator) Semiconductor Transistors", *Solid-State Electronics*, vol. 9, pp. 927–937, 1966.

[23] D. R. Poole, and D. L. Kwong, "Two-dimensional analytical modeling of threshold voltage of short-channel MOSFETs", *IEEE Electron Device Letters*, vol. EDL-5, pp. 443–446, 1984.

[24] D. B. Klaassen, "Compact Modelling of Submicron CMOS", *IEEE Proceedings of the 22nd European Solid-State Circuits Conference (ESSCIRC)*, pp. 4046, Sep. 1996.

[25] M. Miura-Mattausch, U. Feldmann, A. Rahm, M. Bollu, and D. Savignac, "Unified complete MOSFET model for analysis of digital and analog circuits", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 15, pp. 1-7, 1996.

[26] M. Miura-Mattausch, H. Ueno, H. J. Mattausch, K. Morikawa, S. Itoh, A. Kobayashi, and H. Masuda, "100nm-MOSFET Model for Circuit Simulation: Challenges and Solution", *IEICE Transactions on Electronics*, vol. E86-C, no. 6, pp. 1009–1021, Jun. 2003.

[27] A R Boothroyd, S W Tarasewiez, and C Slaby, "MISNAN - a physically based continuous MOSFET model for CAD applications", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, no. 12, pp. 1512–1529, Dec. 1991.

[28] K Joardar, K K Gullapalli, C C McAndrew, M E Burnham, and A Wild, "An improved MOSFET model for circuit simulation", *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 797–801, Jan. 1998.

[29] G. Gildenblat, H Wang, T-L Chen, X. Gu, and X Cai, "SP: An Advanced Surface-Potential-Based Compact MOSFET Model", *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1394–1406, Sep. 2004.

[30] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. J. Scholten, and D. B. M. Klaassen, "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation", *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1979–1993, Sep. 2006.

[31] R Howes, W Redman-White, K G Nichols, P J Mole, M J Robinson, and S Bird, "An SOS MOSFET model based on calculation of the surface potential", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 6, pp. 494–506, Apr. 1994.

[32] M S L Lee, B M Tenbroek, W Redman-White, J Benson, and M Uren, "A physically based compact model of partially depleted SOI MOSFETs for analogue circuit simulation", *IEEE Journal of Solid-State Circuits*, vol. 36, no. 1, pp. 110–121, Jan. 2001.

[33] R van Langvelde, *A Compact MOSFET Model for Distortion Analysis in Analog Circuit Design*, University Press Facilities, Eindhoven, 1998.

[34] D. Suh and J. G. Fossum, "A physical charge-based model for non-fully depleted SOI MOSFETs and its use in assessing floating-body effects in SOI CMOS circuits", *IEEE Transactions on Electron Devices*, vol. 42, no. 4, pp. 728–737, Apr. 1995.

[35] C F Edwards, B M Tenbroek, M S L Lee, W Redman-White, and M J Uren, "The effect of body contact series resistance on SOI CMOS amplifier stages", *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2290–2294, Dec. 1997.

[36] W Redman-White, C Easson, J Benson, R L Rabe, and M J Uren, "A $\Sigma\Delta$ Modulator with Extended Supply Voltages in $0.8\mu m$ SOI CMOS for Direct Ground Referred Instrumentation Interfacing", *IEEE Proceedings of the 25th European Solid-State Circuits Conference (ESSCIRC)*, pp. 322-325, Sep. 1999.

Chapter 2

Low Field DC Model

2.1 Introduction

For a truly *partially depleted* SOI MOSFET, there is no back gate coupling effect [1]. Furthermore, in the case of analogue circuits, the back gate (or substrate) is invariably tied to ground. With a threshold voltage typically of the order of tens of volts, the back device is operating deep within the subthreshold regime and will pass negligible current. This also implies that there is negligible charge at the back interface, as well as a much smaller gate capacitance due to the thicker oxide, leading to negligible impact on transient behaviour. It is therefore assumed that the back device has no significant effect on the total device behaviour. Thus the initial analysis for the STAG model proceeds as for bulk MOSFETs, although as we shall see, the core model presented here will include one additional effect not seen in other treatments. The model described below is for a n-channel device, though a similar analysis can be performed for a p-channel device if the polarity of the voltages and the direction of currents are reversed.

It should be noted that because we are dealing with SOI technology, we will be explicitly referring to the front gate of the device. Thus, for instance, we will talk about C_{of}, the front gate oxide capacitance per unit area. In a bulk device, this would be simply the gate oxide capacitance per unit area (usually denoted by C_{ox}).

2.2 Channel Current

Fig. 2.1 shows a schematic of an SOI MOSFET and the direction of the displacement variables used in this analysis. Using standard assumptions (unipolar device, infinitesimally thin inversion layer, single direction of current flow [2–4]), the charge sheet model expression for a MOSFET may be obtained

$$I_{\mathrm{CH}}(y) = -W\mu_{\mathrm{s}}(y)\; q_{\mathrm{c}}(y)\,\frac{d\psi_{\mathrm{s}}(y)}{dy} + W\mu_{\mathrm{s}}(y)\;\phi_{\mathrm{t}}\frac{dq_{\mathrm{c}}(y)}{dy}. \tag{2.1}$$

where I_{CH} is the channel current, W is the channel width, μ_s is the surface mobility of the carriers, q_c is the channel charge per unit area, ψ_s is the surface potential, and ϕ_{t} is

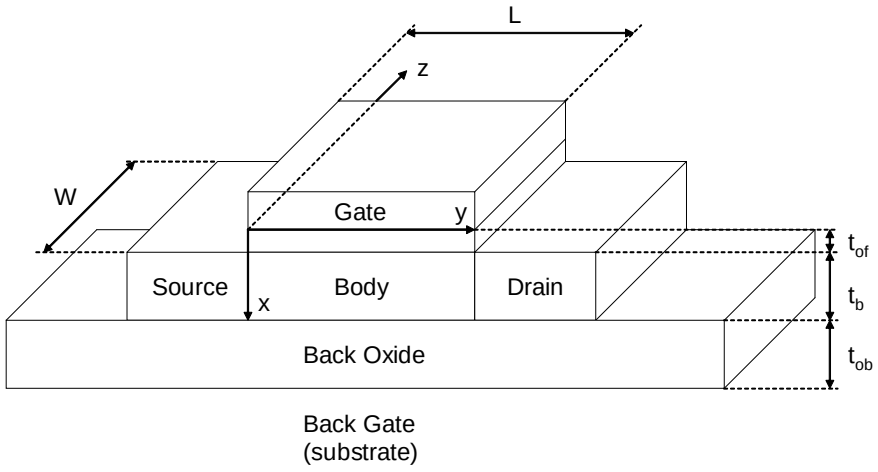

Figure 2.1: *Schematic of Silicon-on-Insulator (SOI) device, showing the coordinate orientation of x, y, and z*

the thermal voltage, given by

$$\phi_t = \frac{k_B T}{q} \tag{2.2}$$

where k_B is Boltzmann's Constant, T is the temperature, and q is the charge on an electron.

Using the electron continuity equation [2,3] and neglecting recombination and generation currents means $\frac{dI_{\rm CH}}{dy} = 0$ for all y, yielding

$$I_{\rm CH} = \frac{W}{L}\Big(-\int_{\psi_{\rm s0}}^{\psi_{\rm sL}} \mu_{\rm s}(y)\; q_{\rm c}(y)\cdot d\psi_{\rm s} + \phi_{\rm t}\int_{q_0}^{q_{\rm L}} \mu_{\rm s}(y)\cdot dq_{\rm c}\Big) \tag{2.3}$$

where L is the channel length, and ψ_{sL} and ψ_{s0} are the surface potentials at the drain and source end of the channel respectively. At this point we adopt the effective mobility approximation [5]. Although the mobility μ_s is a complicated function of a carrier's postion in the channel, we assume for now that μ_s can be represented by a constant value at some average gate and drain field. The dependence of mobility on the transverse and lateral electric fields will be reintroduced into the model in Chapters 3 and 4. Equation (2.3) may now be expressed as [6]

$$I_{\rm CH} = \frac{W}{L}\mu_{\rm s}\Big(-\int_{\psi_{\rm s0}}^{\psi_{\rm sL}} \; q_{\rm c}(y)\cdot d\psi_{\rm s} + \phi_{\rm t}\int_{q_0}^{q_{\rm L}} dq_{\rm c}\Big) \tag{2.4}$$

where q_0 and q_L are the values of the channel charge at the source and drain.

In order to evaluate this expression, we need to do two things. We need to derive an expression for the surface potential ψ_s as a function of the terminal voltages, and we need to define the channel charge q_c in terms of the suface potential. We can then evaulate both quantities at the source and drain. We begin by obtaining an expression for q_c.

2.3 MOS Capacitor Structure

Our treatment of the MOS capacitor will include one element not usually considered [2–4]. We will be assuming the presence of a polysilicon layer, which, under the right conditions, can be depleted close to the polysilicon/gate oxide interface. Figure 2.2 shows the structure under consideration. It is the same starting point used by Arora to formulate his treatment of polysilicon depletion [7]. Arora's method used a surface potential based approach, but his expression for the voltage drop across the depleted polysilicon region was then used to modify the standard threshold voltage expression, a common approach at that time. Indeed, some subsequent model implementations, such as the one by Gildenblat et al [8], have not used Arora's treatment as a starting point because it assumes a fixed surface potential in strong inversion; this assumption is not very accurate. Instead, the approach used in [8], and in most other surface potential models, is to treat the effects of polysilicon depletion as a perturbation to the core surface potential treatment. We will show here that it is entirely appropriate to use Arora's original expression, with suitable modifications, and that furthermore, the effect can be readily incorporated into the core surface potential model itself, without loss of accuracy.

We shall begin by deriving the expression for ψ_p, the voltage drop across the depleted polysilicon region. The treatment proceeds in the same way as in [7]. We begin by applying Gauss' Law and the potential balance equation across the front gate oxide, to give the following relation

$$V_{GfB} = V_{FB} + \eta_s \psi_s(y) + \psi_p(y) - \frac{q_{\rm tot}(y)}{C_{of}} \tag{2.5}$$

where V_{GfB} is the front gate-body voltage, V_{FB} is the front flat band voltage, $q_{\rm tot}$ is the total charge density in the body of the device, C_{of} is the front gate capacitance per unit area, and η_s accounts for the influence of the fast surface states at the silicon-oxide interface as follows

$$\eta_s = 1 + \frac{qD_{itf}}{C_{of}} \tag{2.6}$$

with D_{itf} denoting the fast surface state density, as per the STAG2 model [6].

Equation (2.5) is similar to the standard MOS capacitor expression [2, 3], with an additional term ψ_p denoting the voltage drop across the polysilicon depletion region. A

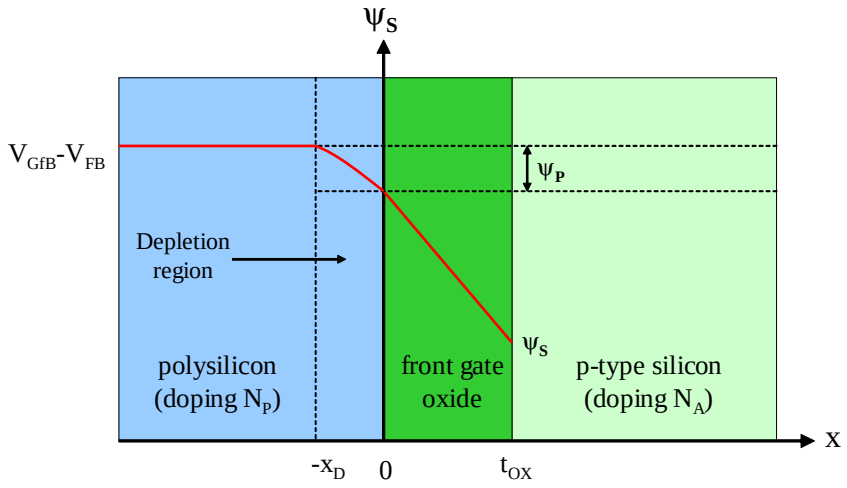

Figure 2.2: *Voltage profile across MOS capacitor structure.*

straightforward application of Poisson's Equation in the depleted polysilicon region gives the standard result for q_p, the charge density in the depletion region.

$$q_p(y) = \gamma_p C_{of} \sqrt{\psi_p(y)} \tag{2.7}$$

with γ_p being the polysilicon body factor, given by

$$\gamma_p = \frac{\sqrt{2q\epsilon_{si}N_P}}{C_{of}} \tag{2.8}$$

where N_P is the doping concentration in the polysilicon layer, and ϵ_{si} is the electrical permittivity of silicon. By assuming zero oxide charge, we can set $q_p(y)$ to $q_{\text{tot}}(y)$ in Equation (2.5). Rearranging then gives a quadratic equation in $\psi_p(y)$

$$\psi_p(y)^2 - 2\left(\omega + \frac{\gamma_p^2}{2}\right)\psi_p(y) + \omega^2 = 0 \tag{2.9}$$

where

$$\omega = V_{GfB} - V_{FB} - \eta_s \psi_s(y) \tag{2.10}$$

By solving this equation, choosing the negative root, and rearranging, we arrive at the following expression for $\psi_p(y)$

$$\psi_p(y) = \left(\sqrt{\omega + \frac{\gamma_p^2}{4}} - \frac{\gamma_p}{2}\right)^2 \tag{2.11}$$

2.4 Channel and Body Charge

The channel charge, consisting of mobile carriers, and the immobile depletion charge, together constitute the total charge density in the active region of the PD SOI MOSFET. This relationship may be expressed as

$$q_{\text{tot}} = q_c + q_b \tag{2.12}$$

where q_c denotes the inversion charge (also known as the channel charge) density, and q_b is the depletion charge (body charge) density. The body charge expression is very straightforward; this is analogous to the bulk charge in a standard MOSFET and (by using the depletion approximation) is generally taken to be [2, 3]

$$q_b(y) = -\gamma_s C_{of} \sqrt{\psi_s(y)} \tag{2.13}$$

with γ_s being the silicon body factor, given by

$$\gamma_s = \frac{\sqrt{2q\epsilon_{si}N_B}}{C_{of}} \tag{2.14}$$

The channel charge expression is slightly more complicated, and deviates from the standard form that is generally used [2, 3], due to the presence of the polysilicon depletion layer. We first substitute Equation (2.11) into Equation (2.5) to obtain the following relation for the total charge density

$$q_{\text{tot}}(y) = -C_{of}\left(\gamma_p\sqrt{\omega + \frac{\gamma_p^2}{4}} - \frac{\gamma_p^2}{2}\right) \tag{2.15}$$

Combining Equations (2.12), (2.13), and (2.15) then yields an expression for q_c

$$q_c(y) = -C_{of}\left(\gamma_p\sqrt{\omega + \frac{\gamma_p^2}{4}} - \frac{\gamma_p^2}{2} - \gamma_s\sqrt{\psi_s(y)}\right) \tag{2.16}$$

We can compare this with the widely-used expression for the channel charge (note that η_s is assumed to be set to unity (zero fast suface state density) in the literature)

$$q_c(y) = -C_{of}\left(\omega - \gamma_s\sqrt{\psi_s(y)}\right) \tag{2.17}$$

Equations (2.13) and (2.16) provide the exact expressions for the body and channel charge. However, it is standard practise [2, 3], to linearise these expressions with respect to the surface potential, in order to make the model more mathematically tractable. It can be seen that under the standard model, the only non-integer power term arises from the square root dependency of the body charge. In the new STAG3 model however, a

second square root term has been introduced into the channel charge due to the addition of the polysilicon layer. Taking a first order Taylor expansion of Equation (2.13) and Equation (2.16) around some arbitary point ψ_{st0}, we obtain the following

$$q_b(y) \approx q_{b0} - C_{of}\delta_s\gamma_s(\psi_s(y) - \psi_{st0}) \tag{2.18}$$

$$q_c(y) \approx q_{c0} - C_{of}\left(\delta_s\gamma_s + \delta_p\gamma_p\right)\left(\psi_s(y) - \psi_{st0}\right) \tag{2.19}$$

where q_{b0} and q_{c0} are $q_b(y)$ and $q_c(y)$ evaluated at the point $\psi_s = \psi_{st0}$

$$q_{b0} = -\gamma_s C_{of}\sqrt{\psi_{st0}} \tag{2.20}$$

$$q_{c0} = -C_{of}\left(\gamma_p\sqrt{\omega + \frac{\gamma_p^2}{4}} - \frac{\gamma_p^2}{2} - \gamma_s\sqrt{\psi_{st0}}\right) \tag{2.21}$$

and δ_s and δ_p are given as

$$\delta_s = \frac{1}{2\sqrt{\psi_{st0}}} \tag{2.22}$$

$$\delta_p = \frac{1}{2\sqrt{\omega + \frac{\gamma_p^2}{4}}} \tag{2.23}$$

We limit ourselves to a first order expansion only, in order to keep the expressions simple when developing the rest of the model. Because of this, Equations (2.22) and (2.23) give poorer accuracy at higher drain voltages. We would therefore like to find a way of improving these expressions so that our linearised charges more closely match the exact values.

Let us first consider δ_s, since this is a well-researched problem, and numerous empirical schemes have been devised to improve matching between the approximate and exact forms for q_c [3]. The problem can be seen in Figure 2.3; the approximated channel charge drops below zero, and is offset from the exact value in the strong inversion regime. We have set N_p, the polysilicon doping concentration, to 10^{24}cm^{-3}, in order to make the influence of polysilicon depletion, and hence the impact of δ_p, negligible. In order to improve the accuracy of δ_s, it has been decided to retain the form used in STAG2 [6]

$$\delta_s = \frac{1}{2\sqrt{1 + \psi_{st0}}} \tag{2.24}$$

Note that this expression is well conditioned even when $\psi_{st0} = 0$. The improvement in accuracy is also shown in Figure 2.3.

We now turn our attention to δ_p. In Figure 2.4, we have plotted the approximated channel charge, but this time using the modified version of δ_s from (2.24). Furthermore, we have set N_p to 10^{19}cm^{-3}, for which appreciable polysilicon depletion will occur. This allows us to evaluate the accuracy of (2.23). It can be seen that the approximated channel charge is

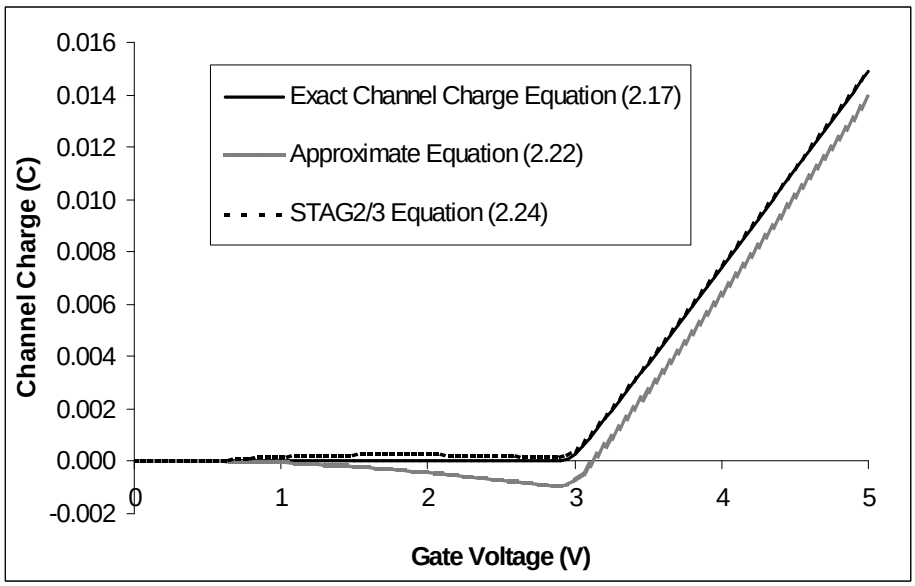

Figure 2.3: *Comparison of two channel charge approximation methods ($V_{DS} = 1.8$)*

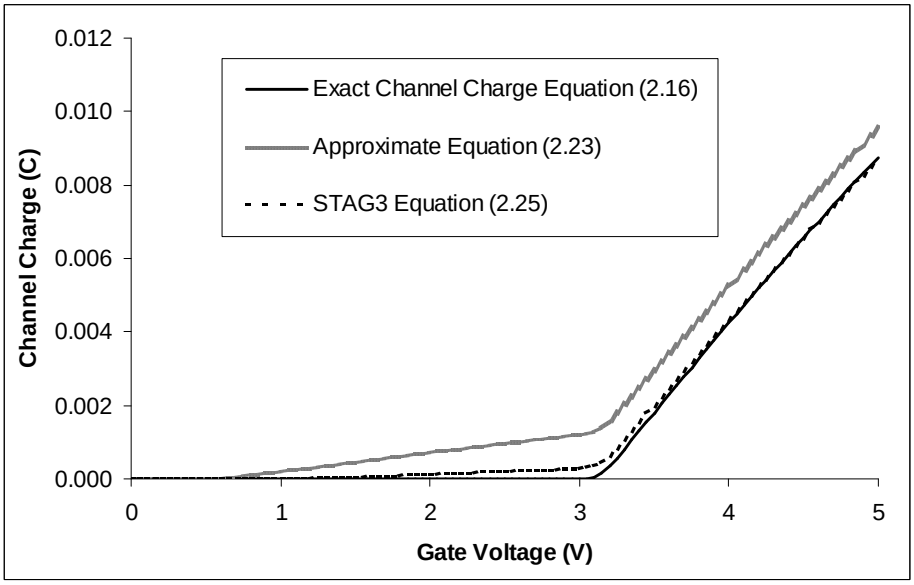

Figure 2.4: *Comparison of two channel charge approximation methods, with polysilicon depletion included ($V_{DS} = 1.8$)*

greatly overestimated. Since no other published compact models incorporate polysilicon depletion directly into the surface potential equations, there is no previous work which addresses this issue. However, it has been found empirically that good matching can be

achieved if the denominator is adjusted in the following way

$$\delta_p = \frac{1}{2\sqrt{\left(1 - M_{sp}\frac{\gamma_s}{\gamma_p}\right)\left(\omega + \frac{\gamma_p^2}{4}\right)}} \tag{2.25}$$

where M_{sp} is an empirical fitting parameter. It has been found that for all the technologies studied in this work, a value of $M_{sp} = 0.8$ gives good results. The improvement can be seen in Figure 2.4. The general form of (2.25) means that it becomes equal to (2.23) in the limit $\gamma_p \gg \gamma_s$ (i.e. for negligible polysilicon depletion). As γ_p approaches γ_s, the empirical factor acts to reduce the denominator, leading to an overall reduction in the channel charge and hence greatly improved accuracy.

Having obtained an accurate approximate expression for the channel charge, we will now write it in the same simple form used in STAG2 [6]

$$q_c(y) = -C_{of}\left[V_{GT} - \alpha\psi_s(y)\right] \tag{2.26}$$

where

$$\alpha = \delta_s\gamma_s + \delta_p\gamma_p \tag{2.27}$$

$$V_{GT} = -\frac{q_{c0}}{C_{of}} + \alpha\psi_{st0} \tag{2.28}$$

Note that V_{GT} and α are defined differently to [6], as a result of adding the polysilicon depletion term.

2.5 Channel Current as a function of Surface Potential

Substituting (2.26) into (2.4) results in

$$I_{CH} = \frac{W}{L}\mu_s C_{of}(f(\psi_{sL}) - f(\psi_{s0})) \tag{2.29}$$

with

$$f(\psi_s) = \left\{V_{GBT} - \frac{\alpha}{2}\psi_s\right\}\psi_s \tag{2.30}$$

where

$$V_{GBT} = V_{GT} + \phi_t\alpha \tag{2.31}$$

All that now remains is for the estimate ψ_{st0} (used for the charge linearisation) and the surface potentials at the source and drain ends (ψ_{s0} and ψ_{sL}) to be found.

2.6 Low Field Surface Potential Model

In most modern circuits, the accumulation region ($V_{GfB} < V_{FB}$) is seldom used. By neglecting this region, the expression for the channel charge obtained by solving the one

dimensional Poisson's Equation (applied across the interface between gate oxide and the body) can be simplified to [2]

$$q_c(y) = -\gamma_s C_{of}\left[\sqrt{\psi_s(y) + \phi_t \exp\left(\frac{\psi_s(y) - 2\phi_F - V_{cb}(y)}{\phi_t}\right)} - \sqrt{\psi_s(y)}\right] \tag{2.32}$$

where ϕ_F is the Fermi potential of the device, and V_{cb} is the channel potential, with $V_{cb}(0) = V_{sb}$ and $V_{cb}(L) = V_{db}$. By setting Equation (2.32) equal to the exact expression for the channel charge, Equation (2.16), and rearranging, we obtain an expression for the surface potential

$$\frac{\gamma_p^2}{\gamma_s^2}\left(\omega(y) + \frac{\gamma_p^2}{2} - \gamma_p\sqrt{\omega(y) + \frac{\gamma_p^2}{4}}\right) = \psi_s(y) + \phi_t \exp\left(\frac{\psi_s(y) - 2\phi_F - V_{cb}(y)}{\phi_t}\right) \tag{2.33}$$

This is an implicit equation, and no exact closed form solution for the surface potential can be derived. Again, because of the inclusion of the polysilicon layer, Equation (2.33) differs from the classic expression frequently cited in the literature [6, 9, 10], given below

$$\left[\frac{V_g - \eta_s\psi_s}{\gamma}\right]^2 = \psi_s(y) + \phi_t \exp\left(\frac{\psi_s(y) - 2\phi_F - V_{cb}(y)}{\phi_t}\right) \tag{2.34}$$

However, the basic problem is the same in both cases; an implicit equation requires time-consuming iterative cycles to solve, and to avoid this we need to employ some method of approximation to obtain a closed-form expression. As was discussed in Chapter 1, not all compact models opt for this approach. Some instead solve the implicit equation iteratively, trading speed for accuracy [9,13,14]. However, one of the key aims when developing STAG3 has been to retain closed form expressions, since simulation of large analogue circuits is very time-intensive. We therefore turn our attention to analytical approximation methods.

Earlier surface potential-based models tended to use empirical techniques to solve the implicit equation. One possibility is to store pre-calculated solutions in a 2-D array and using interpolation for points in between stored values [15]. However, the model requires the partial derivatives of ψ_s (which are needed for the Newton-Raphson technique employed by most circuit simulators) to be derived numerically, thus consuming additional CPU time. Another approach is the approximation of the solution using cubic spline functions [16]. This method is more efficient computationally as the derivatives can be obtained analytically, but it only guarantees continuity of derivatives up to second order.

The use of a function which approximates the true solution to Equation (2.34) has been proposed for standard MOSFETs [17,18]. The technique involves finding asymptotic solutions in the sub-threshold and strong inversion regions and then joining them in a smooth manner. Unfortunately, the proposed function in [18] posesses a discontinuous first derivative, while the function presented in [17] possesses continuity up to the first derivative only.

Subsequent implementations in models such as STAG [6], and the University of Eindhoven model [19] have used infinitely differentiable expressions to avoid this problem. As stated by Chen and Gildenblat [10], these implementations are typically accurate to within 2-3mV. This is certainly acceptable for DC simulation work, but can result in deviations from the iterative solution when calculating transconductances in moderate inversion. To improve the matching, a new analytical method is presented for the SP compact model in [10]. Rather than match two asymptotic approximations, the approach taken here is to approximate the entire solution, resulting in improved accuracy in medium inversion (typical deviations have been found to be of the order of just 10nV).

While the accuracy of the STAG3 low field model could have been improved by using the same approach as [10], the practical benefit was not judged to be worth the effort of implementing a completely new algorithm. While it is one thing to obtain precise matching to the implicit surface potential equation, several other factors come into play when trying to accurately simulate the moderate inversion regime in a real MOSFET device. Of particular importance are high field effects such as mobility degradation (covered in Chapters 3 and 4) and quantum mechanical effects (discussed in Chapter 5). Neither of these types of effect can be modelled with the precision seen in [10], and we can realistically expect the total error from all sources to be of the order of at least a few tens of mV, at least prior to optimisation of model parameters. Any low field deviation in medium inversion of a few mV is therefore likely to be relatively unimportant.

2.7 Surface Potential Determination

We will begin by dividing our treatment into two parts: subthreshold and strong inversion. In each case, we will look at how to simplify Equations (2.33) into an analytical form which is accurate for the region under consideration. Once that has been done for both cases, we will construct a single, continuous, and closed-form expression which gives a good approximation for the surface potential over the whole range of terminal voltages.

2.7.1 Subthreshold

In this region, the channel charge is negligible compared with the body charge, and ψ_s is almost constant. Denoting $\psi_s(y)$ by the constant ψ_{ss}, and setting $q_{tot} \simeq q_b = -\gamma_s C_{of}\sqrt{\psi_{ss}}$, we can obtain a modified version of Equation (2.5) valid under these conditions

$$V_g = \eta_s \psi_{ss} + \psi_p + \gamma_s \sqrt{\psi_{ss}} \tag{2.35}$$

where

$$V_g = V_{GfB} - V_{FB} \tag{2.36}$$

Note that ψ_p is no longer position dependent. We can expand out Equations (2.10) and

(2.11) and obtain a new subthreshold expression for ψ_p

$$\psi_p = V_g - \eta_s \psi_{ss} + \frac{\gamma_p}{2} - \gamma_p \sqrt{V_g - \eta_s \psi_{ss} + \frac{\gamma_p}{4}} \tag{2.37}$$

Substituting Equation (2.37) into Equation (2.36) and rearranging leads to

$$\frac{\gamma_p}{2} + \gamma_s \sqrt{\psi_{ss}} = \gamma_p \sqrt{V_g - \eta_s \psi_{ss} + \frac{\gamma_p}{4}} \tag{2.38}$$

If we now square both sides of Equation (2.38) and factorise, we obtain a quadratic in $\sqrt{\psi_{ss}}$

$$\left[\frac{\gamma_s^2}{\gamma_p^2} + \eta_s\right] \psi_{ss} + \gamma_s \sqrt{\psi_{ss}} - V_g = 0 \tag{2.39}$$

Taking the positive quadratic root, we obtain

$$\psi_{ss} = \left[-\frac{\gamma_s}{2K_{sp}} + \sqrt{\frac{\gamma_s^2}{4K_{sp}^2} + \frac{V_g}{K_{sp}}}\right]^2 \tag{2.40}$$

where

$$K_{sp} = \frac{\gamma_s^2}{\gamma_p^2} + \eta_s \tag{2.41}$$

Note that ψ_{ss} is the saturation potential for the charge sheet model without the linearisation of the body charge. This is because ψ_{ss} is the solution for $q_c = 0$ with $q_b = -\gamma C_{of} \sqrt{\psi_s}$. Problems can occur if the term inside the square root becomes negative, but note that if $V_g < 0$ then Equation (2.33) is not valid (the accumulation condition was specifically excluded in order to simplify the one dimensional solution of Poisson's equation). However, during Newton-Raphson iterations, the gate voltage may venture into this region. Therefore, the following equation is used to ensure that V_g remains positive.

$$V_{gy} = \phi_t \ln\left(1 + \exp\left(\frac{V_g}{\phi_t}\right)\right) \tag{2.42}$$

and V_{gy} is used instead of V_g in Equation (2.40).

2.7.2 Strong Inversion

In strong inversion, the condition $\psi_s > 2\phi_F + V_{cb}$ is met. Looking at Equation (2.33), it can be seen that under this condition, the exponential term is dominant. We can therefore obtain a closed form approximation by setting the other terms of ψ_s in Equations (2.10) and (2.33) equal to some value ϕ_B. We can thus obtain the following expression

$$\frac{\gamma_p^2}{\gamma_s^2}\left(\omega_{sinv} + \frac{\gamma_p^2}{2} - \gamma_p \sqrt{\omega_{sinv} + \frac{\gamma_p^2}{4}}\right) = \phi_B + \phi_t \exp\left(\frac{\psi_{sinv} - 2\phi_F - V_{cb}}{\phi_t}\right) \tag{2.43}$$

and

$$\omega_{sinv} = V_g - \eta_s \phi_B \tag{2.44}$$

Here ψ_{sinv} is the approximated strong inversion potential. A simple rearrangement of Equation (2.43) yields the following result

$$\psi_{sinv} = 2\phi_F + V_{cb} + \phi_t \ln\left\{\frac{1}{\phi_t}\left[\frac{\gamma_p^2}{\gamma_s^2}\left(\omega_{sinv} + \frac{\gamma_p^2}{2} - \gamma_p\sqrt{\omega_{sinv} + \frac{\gamma_p^2}{4}}\right) - \phi_B + \phi_t)\right]\right\} \quad (2.45)$$

where a factor ϕ_t has been added so that $\psi_{ss} = \psi_{si}$ when the following standard threshold condition is met

$$V_{GfB} = V_{FB} + \phi_B + \gamma_s\sqrt{\phi_B} \quad (2.46)$$

Of course, we are still left with the problem of deciding the value to set ϕ_B. The obvious choice, used in STAG2 [6], is to set $\phi_B = 2\phi_F + V_{cb}$. This is a good approximation when looking at the strong inversion region, but there is a loss of accuracy in the moderate inversion region. An alternative expression, proposed by van Langevelde [20], has a functional dependence on V_{GfB}, rather than a constant value. The constant value ϕ_B is replaced by ψ^*, where

$$\psi^* = 2\phi_F + V_{cb} + \frac{\psi_{ss} - 2\phi_F - V_{cb}}{\sqrt{1 + \left(\frac{\psi_{ss} - 2\phi_F - V_{cb}}{4\phi_t}\right)^2}} \quad (2.47)$$

The inclusion of the additional term in Equation (2.47) causes the surface potential in moderate inversion to decrease more gradually, so that it follows the exact iterative solution much more closely.

In [20], it was found that Equation (2.47) need only be applied to one term of the surface potential expression for good accuracy to be obtained. Similarly, it has been found that for the new STAG3 formulation, we need only modify our expression for ω_{si}, by replacing ϕ_B with ψ^*. We can therefore conclude our strong inversion treatment with our final expression for ψ_{sinv}.

$$\psi_{sinv} = 2\phi_F + V_{cb} + \phi_t \ln\left\{\frac{1}{\phi_t}\left[\frac{\gamma_p^2}{\gamma_s^2}\left(\omega_{sinv} + \frac{\gamma_p^2}{2} - \gamma_p\sqrt{\omega_{sinv} + \frac{\gamma_p^2}{4}}\right) - (2\phi_F + V_{cb}) + \phi_t\right]\right\} \quad (2.48)$$

where

$$\omega_{sinv} = V_g - \eta_s\left[2\phi_F + V_{cb} + \frac{\psi_{ss} - 2\phi_F - V_{cb}}{\sqrt{1 + \left(\frac{\psi_{ss} - 2\phi_F - V_{cb}}{4\phi_t}\right)^2}}\right] \quad (2.49)$$

2.7.3 A Single Piece Model

Equations (2.48) and (2.49) provide the basis for our development of a continuous expression for the device surface potential ψ_s. As was done for [20], we would like to define a

function f which gives a smooth transition between the sub-threshold surface potential ψ_{ss} and $2\phi_F + V_{cb}$. In [20], the following expression was used

$$f = 0.5\left[\psi_{ss} + 2\phi_F + V_{cb} - \sqrt{(\psi_{ss} - 2\phi_F - V_{cb})^2 + 4\epsilon^2}\right] \tag{2.50}$$

where ϵ is a smoothing function and is set equal to ϕ_t. However, it was found that using this expression led to poor accuracy in the subthreshold region. In order to correct this, we instead use an exponential transform, similar to those used in STAG2.

$$f = \psi_{ss} + \phi_t \ln\left[\frac{1 + \exp\left[\frac{-(2\phi_F + V_{cb})}{\phi_t}\right]}{1 + \exp\left[\frac{\psi_{ss} - (2\phi_F + V_{cb})}{\phi_t}\right]}\right] \tag{2.51}$$

The exponential helps ensure that f becomes much closer to ψ_{ss} in subthreshold, compared to Equation (2.50). The next step is to simply replace all occurrences of $2\phi_F + V_{cb}$ with f in Equations (2.48) and (2.49).

$$\psi_{st} = f + \phi_t \ln\left\{\frac{1}{\phi_t}\left[\frac{\gamma_p^2}{\gamma_s^2}\left(\omega + \frac{\gamma_p^2}{2} - \gamma_p\sqrt{\omega + \frac{\gamma_p^2}{4}}\right) - f + \phi_t\right]\right\} \tag{2.52}$$

where

$$\omega = V_g - \eta_s\left[f + \frac{\psi_{ss} - f}{\sqrt{1 + \left(\frac{\psi_{ss} - f}{4\phi_t}\right)^2}}\right] \tag{2.53}$$

However, it was found that yet another transform needed to be applied in order to get acceptable subthreshold characteristics. We therefore introduce f_{trans}

$$f_{trans} = \frac{1}{1 + \exp\left[\frac{\psi_{ss} - (2\phi_F + V_{cb} - 6\phi_t)}{6\phi_t}\right]} \tag{2.54}$$

We apply this to the logarithmic term in Equations (2.52), since this is associated with the strong inversion region, and we would this term to quickly become negligible in subthreshold. The $6\phi_t$ in the numerator and denominator of the exponent ensures a smooth transition in the region just below strong inversion. Thus we obtain our final surface potential expression

$$\psi_{st} = f + (1 - f_{trans})\phi_t \ln\left\{\frac{1}{\phi_t}\left[\frac{\gamma_p^2}{\gamma_s^2}\left(\omega + \frac{\gamma_p^2}{2} - \gamma_p\sqrt{\omega + \frac{\gamma_p^2}{4}}\right) - f + \phi_t\right]\right\} \tag{2.55}$$

Equation (2.55) is valid for long devices where the effects of velocity saturation can be neglected. For the source end, this equation provides a very good estimate of the true surface potential, and so it is used for the estimate ψ_{st0}, i.e.

$$\psi_{st0} = \psi_{st}(V_{GfB}, V_{SB}) \tag{2.56}$$

Similarly, we can obtain an expression for the drain end

$$\psi_{stL} = \psi_{st}(V_{GfB}, V_{DB}) \tag{2.57}$$

Having reached this point, we have now developed a complete surface potential model for the low field case. High field effects such as quantum effects and mobility degradation complicate the picture considerably, and these will be looked at in detail over the next three chapters.

2.7.4 Intrinsic Channel Current Expression

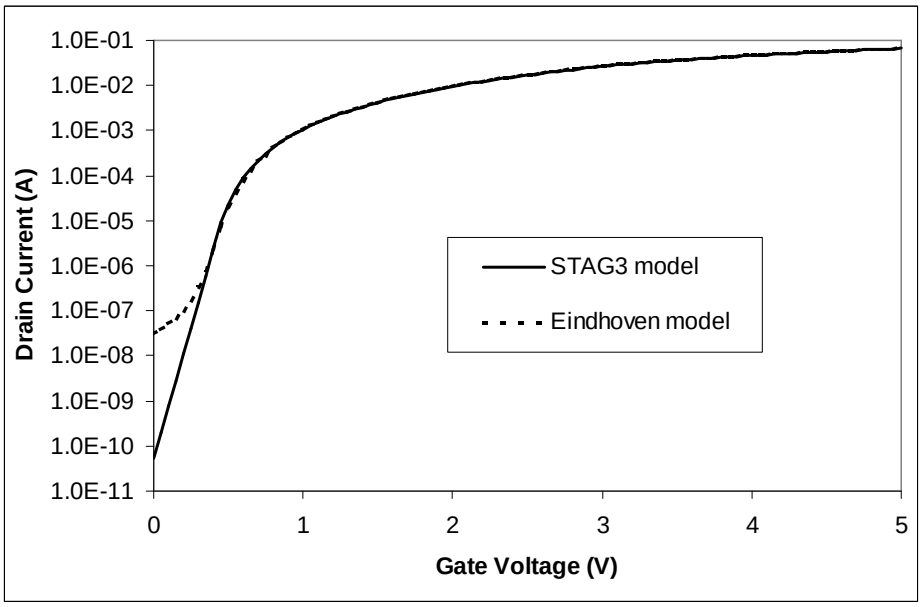

Figure 2.5: *Comparison of subthreshold characteristics using STAG3 and Eindhoven models. It can be seen that the Eindhoven model gives a non-physical subthreshold slope.*

It only remains now to complete our development of the intrinsic channel current expression. Looking back at Section 2.5, we can see that it is simply a case of substituting Equation (2.56) and Equation (2.57) into Equation (2.29) to obtain an intrinsic drain current equation without high field effects

$$I_{CHint} = \frac{W}{L}\mu_0 C_{of}(f(\psi_{stL}) - f(\psi_{st0})) \tag{2.58}$$

where μ_s has been replaced by μ_0, the low field mobility.

In Figure 2.5 we plot the channel current logarithmically as a function of gate voltage, for a drain voltage of $0.1V$. We have also plotted the equivalent result for the Eindhoven

model, which is obtained by replacing Equation (2.51) with Equation (2.50), and setting $f_{trans} = 0$. It can be seen that the new STAG3 expression exhibits more physical behaviour in the subthreshold region.

References

[1] H K Lim and J G Fossum, "Current-voltage characteristics of thin-film SOI MOSFETs in strong inversion", *IEEE Transactions on Electron Devices*, vol. 31, no. 4, pp. 401–408, Apr. 1984.

[2] Y P Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill Book Company, 1987.

[3] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[4] M S L Lee, *Compact Modelling of Partially Depleted Silicon-On-Insulator MOSFETs for Analogue Circuit Simulation*, PhD thesis, University of Southampton, Southampton, UK, Dec. 1997.

[5] R. F. Peirret, and T. A. Shields, "Simplified Long-Channel MOSFET Theory", *Solid-State Electronics*, vol. 26, pp. 143–147, 1983.

[6] M S L Lee, B M Tenbroek, W Redman-White, J Benson, and M Uren, "A physically based compact model of partially depleted SOI MOSFETs for analogue circuit simulation", *IEEE Journal of Solid-State Circuits*, vol. 36, no. 1, pp. 110–121, Jan. 2001.

[7] N D Arora, R Rios, and C L Huang, "Modeling the polysilicon depletion effect and its impact on submicrometer CMOS circut performance", *IEEE Transactions on Electron Devices*, vol. 42, no. 5, pp. 935–943, May 1995.

[8] G Gildenblat, T L Chen, and P Bendix, "Analytical approximation for perturbation of MOSFET surface potential by polysilicon depletion layers", *Electronics Letters*, vol. 35, no. 22, pp. 1974–1976, October 1999.

[9] A R Boothroyd, S W Tarasewiez, and C Slaby, "MISNAN - a physically based continuous MOSFET model for CAD applications", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, no. 12, pp. 1512–1529, Dec. 1991.

[10] T. L. Chen, and G. Gildenblat, "Analytical approximation for the MOSFET surface potential", *Solid-State Electronics*, vol. 45, pp. 335–339, 2001.

[11] N D Arora, R Rios, C L Huang, and K Raol, "PCIM: A physically based continuous short-channel IGFET model for circuit simulation", *IEEE Transactions on Electron Devices*, vol. 41, no. 6, pp. 988–997, June 1994.

[12] H C de Graaf and F M Klaassen, *Compact Transistor Modelling for Circuit Design*, Computational Microelectronics. Springer-Verlag, Wien, New York, 1990.

[13] M. Miura-Mattausch, H. J. Mattausch, H. J. Arora. and C. Y. Yang, "MOSFET Modeling Gets Physical", *IEEE Circuit and Devices Magazine*, vol. 17, no. 6, pp. 29–36, Nov. 2001.

[14] K Joardar, K K Gullapalli, C C McAndrew, M E Burnham, and A Wild, "An improved MOSFET model for circuit simulation", *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 797–801, Apr. 1998.

[15] S. Yu, A. F. Franz, and T. G. Mihran, "A physical parametric transistor model for CMOS circuit simulation", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-7, pp. 1038–1052, 1988.

[16] P. K. K. H. J. Park, and C. Hu, "A charge sheet capactiance model for short channel MOSFET's for SPICE", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-10, pp. 376–389, 1991.

[17] R Howes, W Redman-White, K G Nichols, P J Mole, M J Robinson, and S Bird, "An SOS MOSFET model based on calculation of the surface potential", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 6, pp. 494–506, Apr. 1994.

[18] C. Turchetti, and G. Massetti, "A CAD-Oriented Analytical MOSFET Model for High-Accuracy Applications", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. CAD-3, pp. 117–120, 1984.

[19] Ronald van Langevelde, and Francois M. Klaasen, "Effect of Gate-Field Dependent Mobility Degradation on Distortion Analysis in MOSFET's", *IEEE Transactions on Electron Devices*, vol. 44, no. 11, pp. 2044–2052, 1997.

[20] R van Langvelde, *A Compact MOSFET Model for Distortion Analysis in Analog Circuit Design*, University Press Facilities, Eindhoven, 1998.

Chapter 3

High Field Effects 1: Mobility Model

3.1 Introduction

Probably the single most important development in the understanding of carrier mobility in silicon inversion layers came in 1979, due to the work of Sabnis and Clemens [1]. They showed that plotting carrier mobility versus effective vertical field resulted in a single universal curve at high electric fields, independent of impurity concentration or substrate biasing. Deviations from this universal curve are seen to occur at low fields.

A great deal of work has been done to improve understanding of the physics of inversion layer mobility in MOS structures [2–6]. In particular, there is much interest within the numerical simulation community in developing more accurate mobility models [7–10]. Some of this work is too mathematically complex to be reduced into a convenient form for compact models, but it is useful to understand the main mechanisms behind mobility degradation. It is widely accepted [2, 3, 11] that there are three main scattering mechanisms which impose a vertical electric field dependence on the inversion carrier mobility. These are: phonon scattering, surface roughness scattering, and Coulomb (ionised impurity) scattering. Let us briefly examine each in turn.

Surface phonon scattering can be considered as making a separate contribution to that of the acoustic phonons associated with scattering in bulk carrier transport. This scattering mechanism has been theoretically calculated as having a dependence on the inversion charge q_c with a factor of proportionality of 1/3, provided that the inversion charge dominates over the body charge (i.e. $q_c \gg q_b$) [5]. Since this condition corresponds to strong inversion, and since q_c is proportional to the *effective vertical field* E_{xeff} in this operating regime, this translates to a 1/3 power dependence of the surface phonon mobility degradation as a function of the electric field. This relation has been subsequently confirmed experimentally [2]. This mechanism is more evident in older device technologies, where Coulomb scattering is relatively unimportant, allowing phonon scattering to dominate at lower gate voltages. As the vertical effective field increases, surface roughness starts to

take over, since it has a stronger field dependence.

Surface roughness scattering is the result of carriers scattering from imperfections in the Si/SiO_2 interface, and typically becomes dominant at electric fields in excess of 0.5MV/cm [9]. It has been found experimentally to have approximately a linear dependence on the transverse electric field for holes, and a square dependence for electrons [2], although the exact relation is dependent on the quality of the oxide interface [9]. In modern deep sub-micron devices, the required field strength is present almost immediately above threshold, so that surface roughness becomes important across the whole gate bias range. We would therefore expect phonon scattering to have a relatively minor impact on the characteristics of deep sub-micron devices.

Coulomb scattering is due to ionised impurity atoms in the channel region and oxide. It is more evident in highly doped devices, since these have more ionised impurities to cause scattering events. Furthermore, the effect is most pronounced in the region below threshold, when the inversion charge density is too small to screen the ionised impurities. Devices whose carriers undergo strong Coulomb scattering can be expected to exhibit large mobility degradation at low gate voltages. Mobility degradation due to Coulomb scattering has been found empirically to have an inverse square dependence on the effective electric field [2].

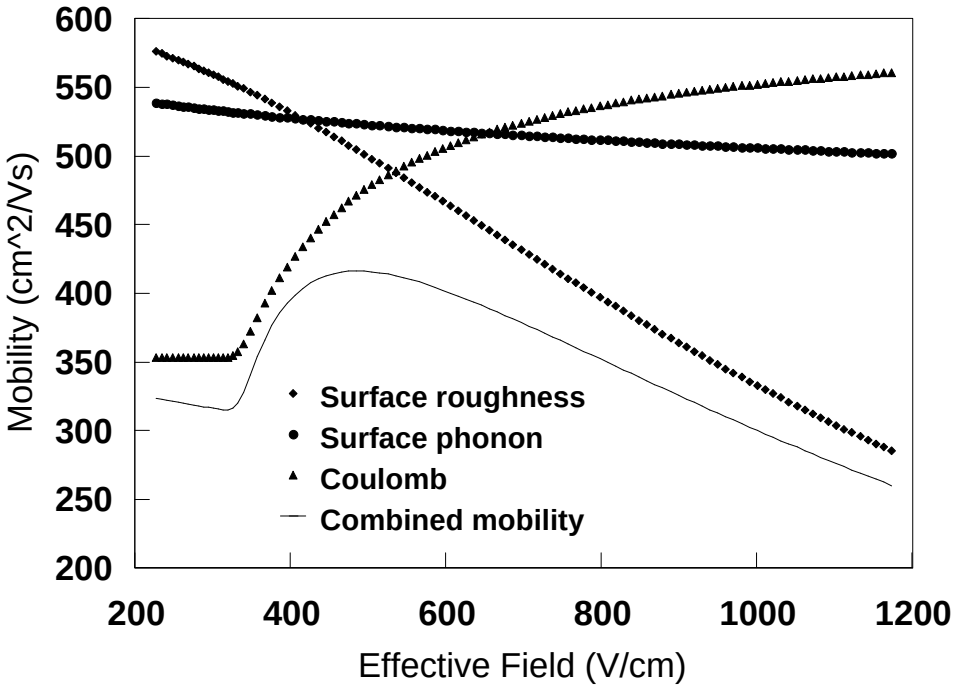

Figure 3.1: *Graph showing functional dependencies of different scattering mechanisms on effective vertical electric field*

The combined effect of these different scattering mechanisms on the mobility is shown in Figure 3.1. It can be seen that each mechanism has more influence in certain regions of device operation, and we can also expect the technology generation to be a factor as well. In older device technologies, we would expect phonon scattering to dominate at low gate voltages, before gradually giving way to surface roughness scattering. In more highly doped deep sub-micron devices, we would expect surface roughness scattering to be evident at or close to threshold, where it would compete with Coulomb scattering. At higher gate voltages, surface roughness would be the dominant mechansim. Any effects of phonon scattering would be likely to be confined to a small region of operation just as strong inversion was reached, so that Coulomb scattering centres would be screened by free carriers, while surface roughness wouldn't yet dominate over the other mechanisms.

3.2 Compact Mobility Models

Having established that inversion carrier mobility is a complicated function of the vertical field, the first issue that arises is how to mathematically describe the average effective vertical field, E_{xeff}, as seen by the carriers in the inversion channel. A commonly used expression is

$$E_{xeff}(\psi_s) = \frac{f_c\, q_c(\psi_s) + f_b\, q_b(\psi_s)}{\epsilon_{Si}} \tag{3.1}$$

Note that we have indicated that E_{xeff} is a function of surface potential, and hence of channel position, due to its dependence on the charges q_c and q_b. It is common practise for compact models to set f_b and f_c to constant values; f_b to 1, and f_c to 0.5 for electrons and 0.33 for holes [12–15]. The real situation is somewhat more complicated. In one frequently cited study, Takagi et al. [16] showed experimentally that f_c should only be equal to 0.5 for electrons moving in the <100> direction. In the <110> and <111> it is instead appropriate to use a value of 0.33.

A theoretical treatment by Krutsick and White [17], showed that it is more appropriate to write E_{xeff} as

$$E_{xeff}(\psi_s) = \frac{f_c q_c(\psi_s) + \left(1 - \frac{\Delta z_I}{x_d}\right) q_b(\psi_s)}{\epsilon_{Si}} \tag{3.2}$$

where Δz_I is the inversion layer centroid (the average distance of the carriers from the Si/SiO_2 interface), and x_d is the depletion layer width. This accounts for the fact that quantum effects cause the inversion layer to move further from the interface, once the transverse electric field becomes sufficiently strong. In the classical low-field/high temperature limit, $\Delta z_I \to 0$, in which case Equation (3.2) becomes equivalent to Equation (3.1) in the case $f_b = 1$. This is discussed in further detail in Chapter 5.

This early work was taken further by Vasileska, Ferry et al. [7]. They were able to show a dependence of f_c and f_b on the level of impurity doping, and went on to conclude that the doping profile could have an additional effect on f_b.

The second issue to be addressed is to find an appropriate expression relating mobility to the field. At one time, it was commonplace for circuit simuators to rely on the classic empirical relation [18, 19]

$$\mu_{xeff} = \frac{\mu_0}{1 + \alpha_\theta E_{xeff}} \tag{3.3}$$

where μ_0 is the bulk mobility, μ_{xeff} is the vertical field degraded effective mobility, and α_θ is a fitting constant. This semi-empirical relation is a simple linear expression, and so provides a high degree of mathematical tractability when formulating model equations. Having a simple mobility relation is helpful when formulating a compact model, since it is highly desirable to have closed form expressions, in order to avoid undesirable iterative algorithms. However, this convenience comes at the expense of accuracy, and while it has been found to give reasonable results for longer channel devices (above 1μm), the model becomes increasingly inaccurate as devices are scaled down into the sub-micron region.

Work undertaken during the course of this study has shown that the standard mobility model, in conjunction with established extraction techniques [19,20], can lead to differences of up to 50% between extracted and optimised values of the mobility model parameters. Even after optimisation, it was difficult to obtain adequate curve fits across the full range of device geometries and bias conditions, using a single parameter set. Indeed, it has often been found necessary to extend parameter optimisation beyond those model parameters directly associated with the mobility expressions. For this reason, the inclusion of various second order short channel effects can improve the fit, whilst at the same time compensating for (and thus concealing) shortcomings in the mobility model. Needless to say, this is something which is best kept to a minimum in a physical model. Second-order effects can always be accounted for empirically, but if they are added to a sound core model, then there is a reduced likelihood of overlap between the different parts of the model and different groups of parameters.

Because of the mathematical convenience of using Equation (3.3), some channel compact models have been content to use what really amounts to a long channel mobility expression [14,21]. Other models have employed fitting parameters to improve the accuracy; the Motorola model [22] and the BSIM3 model [23] are two examples. While this is an improvement, such expressions still try to lump contributions from several scattering mechanisms together, and so they still lack a physical basis. It needs to be appreciated that mobility degradation is a first-order effect, with a major impact on device characteristics, and so any short-cuts taken in simulating this aspect of device behaviour can potentially have

serious repercussions for the overall usefulness of the model.

Although expressions have been devised which incorporate the main physical mobility scattering mechanisms [2, 24], it was not until quite recently that attempts started to be made to incorporate more physical models in circuit simulators [12, 13]. The next logical step-up from the inverse linear relation is to add an extra term to account for the strong contribution of surface roughness scattering at high electric fields. This can be accounted for quite easily by making the denominator a quadratic expression.

$$\mu_{xeff} = \frac{\mu_0}{1 + \theta_1 E_{xeff} + \theta_2 E_{xeff}^2} \tag{3.4}$$

This expression is used in several models [25, 26]. It is still essentially an empirical approach, since θ_1 and θ_2 are treated as fitting parameters and no attempt is made to relate them to the physical scattering mechanisms. One of these models, LETISOI [26], has one further term in the denominator which is proportional to the depletion charge at threshold, and which therefore provides a measure of the contribution from Coulomb scattering. However, this term is constant and thus does not account for the charge screening by the inversion layer. In fact, one consistent trend among all but the most recent generation of compact models is a tendency to attach little importance to the Coulomb scattering contribution. This can lead to errors around threshold, where the Coulomb scattering term, while not necessarily dominant, is certainly non-negligible (see Figure 3.1).

A compact MOSFET model by the University of Eindhoven, designed specifically for accurate modelling of distortion analysis, [12] was among the first compact models to feature a mobility model that explicitly included some of the physics of inversion layer carrier scattering. Surface roughness and phonon scattering mechanisms are included, but Coulomb scattering is not. This has probably been omitted to make the equations easier to solve, and in any case is less important to them, since their model concentrates more on accuracy in strong inversion. Nevertheless, the model is quite physical as a result of these inclusions, especially since additional empirical measures are taken to improve the fit to experimental data [27].

More recent models have included more sophisticated expressions for the mobility degradation. The HiSIM model includes the three scattering components described in this chapter [28]. The PSP model combines the surface roughness and phonon scattering components into a single adjustable power term, together with a separate Coulomb scattering term [29].

3.3 STAG Mobility Model

The new work with the STAG3 model is intended to make certain key improvements to the original STAG model. The old linear model used in STAG2, described in Equation (3.3), has been replaced by a physical mobility model, one which includes a functionally correct Coulomb scattering term, in addition to the phonon and surface roughness contributions. The basic intention is that the new STAG3 mobility expression could be used to reconstruct a complete mobility curve, including both the universal and non-universal components depicted in Figure 3.1.

In order to incorporate the new mobility model into the surface potential treatment introduced in Chapter 2, it is necessary to define some new quantities. Since it would be too complicated to try and model the mobility degradation as a function of channel position, we will take the standard approach of applying an averaged value along the entire channel. We therefore start by obtaining expressions for the channel and body charge, averaged along the channel.

$$q_{bav} = \frac{1}{\psi_{sL} - \psi_{s0}} \int_{\psi_{s0}}^{\psi_{sL}} q_b(\psi_s) d\psi_s \tag{3.5}$$

$$q_{cav} = \frac{1}{\psi_{sL} - \psi_{s0}} \int_{\psi_{s0}}^{\psi_{sL}} q_c(\psi_s) d\psi_s \tag{3.6}$$

Substituting Equation (2.18) into Equation (3.5) and Equation (2.26) into Equation (3.6)

$$q_{bav} = \gamma_s C_{of} \left(\sqrt{\psi_{st0}} + \frac{\delta_s}{2} \left[\psi_{sL} - \psi_{s0} \right] \right) \tag{3.7}$$

$$q_{cav} = C_{of} \left(V_{GT} - \frac{\alpha}{2} \left[\psi_{sL} + \psi_{s0} \right] \right) \tag{3.8}$$

Finally, we can use Equation (3.1) to define the averaged vertical electric field in terms of q_{bav} and q_{cav}.

$$E_{xav} = \frac{(f_c q_{cav} + f_b q_{bav})}{\epsilon_{si}} \tag{3.9}$$

We can now use these averaged values to relate the contribution of each scattering mechansim directly to the surface potential at the source and drain. We shall now examine the form of each scattering term. In the following sections, α is used to denote a model parameter.

3.3.1 Phonon Scattering

The generally accepted dependence of phonon scattering on vertical field in a MOS inversion layer is given by [2].

$$\mu_{ph} \propto E_{xeff}^{-1/3} \tag{3.10}$$

According to [30], it is appropriate to use an inverse linear temperature dependence for μ_{ph}.

$$\mu_{ph} \propto T^{-1} \tag{3.11}$$

The phonon contribution can thus be modelled as

$$G_{ph} = \alpha_{ph} E_{xav}^{1/3} \left[1 + \frac{\Delta T}{T}\right] \tag{3.12}$$

3.3.2 Surface Roughness Scattering

Surface roughness scattering starts to dominate over phonon scattering at higher gate voltages. While the power dependence of this scattering mechanism is somewhat influenced by the interface quality and gate material [9], the most commonly used expression for electron scattering at the oxide interface is

$$\mu_{sr} \propto E_{xeff}^{-2} \tag{3.13}$$

Although [9] found that μ_{sr} should vary with temperature, the dependence is expected to be very weak. We will therefore neglect the temperature term.

$$G_{sr} = \alpha_{sr} E_{xav}^{2} \tag{3.14}$$

3.3.3 Coulomb Scattering

This scattering contribution is most dominant close to the threshold voltage, since scattering due to ionised impurities is greatly reduced by the screening of the inversion layer at higher gate voltages. It has been found empirically that the magnitude of the scattering term is inversely proportional to the channel charge density [2]. In order to obtain a more numerically stable expression, the approach taken by Villa et al [8] was adopted. Since the emphasis is on deriving an accurate expression around threshold, at low inversion charge densities, a non-degenerate version of (10) in [8] was derived. This then leads to an expression of the form:

$$\mu_{cou} = \frac{\mu^*}{N_B \cdot L_{th}} \cdot F^3 \cdot \left(1 + \frac{L_{th}}{F \cdot L_s}\right)^2 \tag{3.15}$$

Some explanation of the terms in this equation is required before we go any further. Firstly, as can be seen by considering the units of the expression, μ^* is actually the unscreened mobility *per scattering centre per unit area*. The Coulomb scattering limited mobility μ_{cou} is inversely proportional to the impurity doping concentration (which we take to be the body doping concentration N_B), and also L_{th}, which is the characteristic "thermal length" of the carriers for a given average thermal energy k_BT. This is treated as a constant, and at room temperatures, $L_{th} = 25\mathring{A}$ [8]. The term in brackets represents the increase in mobility due to screening of scattering centres by the inversion layer, with L_s being a characteristic screening length.

The F^3 term is included to account for the effects of degeneracy. F is defined as ratio between the *Debye-Hückel length* L_{DH} and the effective screening length.

$$F = \frac{L_s}{L_{DH}} \tag{3.16}$$

The reason for there being a cubic term in (3.15) is that the degeneracy has a dual effect on the mobility. Firstly, it increases the average electron momentum. Consequently, the thermal length L_{th}, which is the inverse of the electron wavevector, is decreased by a factor F. Secondly, it also increases the electron kinetic energy, from k_BT to F^2K_BT , so that μ^* is also increased by the same factor.

However, we need to simplify the expression given in (3.15). To do this, we need to consider that the main reason for including the Coulomb scattering model in the first place is so as to improve the accuracy of the mobility model close to the threshold region, since it is only here that it is comparable to the other scattering contributions. In other words, we are only interested in the non-degenerate case, and are less concerned with accuracy further into the strong inversion region, where degeneracy might occur. With this in mind, we can very simply obtain a non-degenerate version of (3.15) by setting $F = 1$.

$$\mu_{cou} = \frac{\mu^*}{N_B \cdot L_{th}} \cdot \left(1 + \frac{L_{th}}{L_s}\right)^2 \tag{3.17}$$

We can also apply the non-degenerate condition to L_s, which from equation [8] gives

$$L_s = L_{DH} = \frac{2\epsilon_{si}kT}{q^2N_c} \tag{3.18}$$

Since $q_c = qN_c$, we can substitute Equation (3.18) into Equation (3.17)

$$\mu_{cou} = \frac{\mu^*}{N_B \cdot L_{th}} \cdot \left(1 + \frac{L_{th} q q_c}{2\epsilon_{si} kT}\right)^2 \tag{3.19}$$

$$\mu_{cou} = \frac{\mu^*}{N_B \cdot L_{th}} \cdot \left(1 + \frac{q_c}{q_{s0}}\right)^2 \tag{3.20}$$

$$\mu_{cou} \propto \left(1 + \frac{q_c}{q_{s0}}\right)^2 \tag{3.21}$$

where q_{s0} is some characteristic charge density associated with the screening of the inversion layer.

$$\mu_{cou} \propto T \tag{3.22}$$

$$G_{cou} = \alpha_{cou} \frac{N_B}{\left[1 + \frac{\Delta T}{T}\right]} \left(\frac{q_{s0}}{q_{s0} + q_c}\right)^2 \tag{3.23}$$

3.3.4 Complete Model

Finally, we need to combine the effects of the different scattering mechanisms in such a way that we are left with an averaged value for the channel mobility, expressed as a function of the electric field (and hence of the surface potential). By far the most common approach is to use Matthiessen's rule [2, 3, 8, 11–13, 16, 24, 30–36], which gives an expression of the form

$$\mu_{xeff} = \left[\frac{1}{\mu_{ph}} + \frac{1}{\mu_{sr}} + \frac{1}{\mu_{cou}}\right]^{-1} \tag{3.24}$$

While the application of Matthiessen's rule to this problem is certainly mathematically convenient, the validity of doing so isn't straightforward from a device physics point of view. The rule is only valid in the case where scattering occurs through short-range, local events. In such a case, each scattering mechanism is unaffected by the others, and the total scattering rate can be expressed as the sum of the individual scattering rates from each mechanism. However, if a given scattering mechanism has a long-range component, this can be influenced by scattering events due to other mechanisms, in which case the different mechanisms cannot be treated separately. In this case, the mechanisms are considered to be coupled, and using Matthiessen's rule to combine their scattering contributions becomes invalid.

A few studies have examined the degree to which Matthiessen's rule can be legitimately applied to the problem of determining carrier mobilities in silicon inversion layers. The first in-depth study was conducted by Lee et al [3], and examined the relation between surface roughness and phonon scattering mechanisms (at the time of the study, Coulomb scattering was less relevant, since channel doping was not as high as today). It was found that Matthiessen's rule could be used to accurately combine the contributions from these two mechanisms, provided that the effects of inter-valley scattering were also accounted for. The occupancy of these different energy valleys is a complicated function of the electric field, but the main trends can be qualitatively described. At low vertical electric fields, the carrier population is fairly equally distributed among all the available energy valleys, which include longitudinal valleys (high effective mass and hence low mobility) and transverse energy valleys (low effective mass, high mobility). At sufficiently high fields, the population redistributes into the high mobility transverse valleys; of course this increase in the mean carrier mobility must be set against the increase in phonon and surface roughness scattering.

The basic upshot of this is that Matthiessen's rule seems to be valid way of combining the scattering contributions due to surface roughness and surface phonons. The fact that account must also be taken of energy valley populations is a separate issue; it doesn't say anything about the interaction between different scattering processes, but rather indicates that there is an additional functional dependence between mobility and electric field - beyond that introduced by each process - which needs to be taken into account in an accurate physics-based model. It is however rare for a compact model to address this issue. To the author's knowledge, only the University of Eindhoven compact model has explicitly acknowledged the findings of [3], attempting to account for this effect through an empirical modification to the standard Matthiessen's rule expression [12].

A second, more recent study by Ishihara and Sano [37] provides a detailed theoretical treatment examining the coupling between the phonon and Coulomb scattering mechanisms. Among other findings, it was concluded that it is more valid to apply Matthiessen's rule in some regions of device operation than others. Specifically, when the inversion charge is relatively high, it is expected that the high carrier charge concentration will screen out the effects of phonon scattering. As a result the two scattering mechanisms will be uncoupled, and Mattiessen's rule is valid. In contrast, for low charge concentrations, there will be a coupling between phonon scattering and Coulomb scattering, such that the Coulomb potential is effectively cut off by phonon scattering. This is unfortunate, since in strong inversion surface roughness is the dominant scattering mechanism, and any errors in the Coulomb and phonon contributions would be less important anyway. Instead, using Matthiessen's rule means that the error will be greatest at low gate fields, which is exactly where we would like an accurate prediction for these two scattering mechanisms.

In summary, there are two factors which might introduce errors when using Equation (3.24),

assuming that our expressions for each scattering contribution are accurate. The first is that coupling between scattering processes might render Matthiesssen's rule invalid. There have only been a limited number of studies in the literature, so it is very difficult to obtain a complete overview on this, but we have seen that we can certainly expect this for Coulomb and phonon scattering at low gate fields. The second factor is the effect of energy valley re-population as we move from low to high gate fields.

The vertical field limited effective mobility μ_{xeff} is given by

$$\mu_{xeff} = \frac{\mu_0}{1 + G_{ph} + G_{sr} + G_{cou}} = \frac{\mu_0}{G_v} \tag{3.25}$$

where G_{ph}, G_{sr}, and G_{cou} are the contributions due to phonon, surface roughness, and Coulomb scattering respectively.

References

[1] A. G. Sabnis, and J. T. Clemens "Characterization of the electron mobility in the inverted <100> Si surfaces", in *IEDM Technical Digest*, 1979, pp. 18–21.

[2] Shin-ichi Takagi, Akira Toriumi, Masao Iwase, and Hiroyuki Tango, "On the Universality of Inversion Layer Mobility in Si MOSFET's: Part I - Effects of Substrate Impurity Concentration", *IEEE Transactions on Electron Devices*, vol. 41, no. 12, pp. 2357–2362, 1994.

[3] Kwyro Lee, Joo-Sun Choi, Sang-Pil Sim, and Choong-Ki Kim, "Physical Understanding of Low-Field Carrier Mobility in Silicon MOSFET Inversion Layer", *IEEE Transactions on Electron Devices*, vol. 38, no. 8, pp. 1905–1912, 1991.

[4] Mou-Shiung Lin, "A Better Understanding of the Channel Mobility of Si MOSFET's Based on the Physics of Quantized Subbands", *IEEE Transactions on Electron Devices*, vol. 35, no. 12, pp. 2406–2411, 1988.

[5] S. Kawaji, "The Two-Dimensional Lattice Scattering Mobility in a Semiconductor Inversion Layer", *Journal of the Physical Society of Japan*, vol. 27, no. 4, pp. 906–908, Oct. 1969.

[6] Sundaram Padmanabhan, and Allen Rothwarf, "Quantum Inversion Layer Mobility: Numerical Results", *IEEE Transactions on Electron Devices*, vol. 36, no. 11, pp. 2557–2566, 1989.

[7] Dragica Vasileska, and David K. Ferry, "Scaled Silicon MOSFET's: Universal Mobility Behaviour", *IEEE Transactions on Electron Devices*, vol. 44, no. 4, pp. 577–583, 1997.

[8] Stefano Villa, Andrea L. Lacaita, Laura M. Perron, and Roberto Bez, "A Physically-Based Model of the Effective Mobility in Highly-Doped n-MOSFET's", *IEEE Transactions on Electron Devices*, vol. 45, no. 1, pp. 110–115, 1998.

[9] Gianluca Mazzoni, Andrea L. Lacaita, Laura M. Perron, and Agostino Pirovano, "On Surface Roughness-Limited Mobility in Highly-Doped n-MOSFET's", *IEEE Transactions on Electron Devices*, vol. 46, no. 7, pp. 1423–1428, 1999.

[10] Dragica Vasileska, Paolo Bordone, Terry Eldridge, and David K. Ferry, "Quantum Transport Calculations for Silicon Inversion Layers in MOS Structures", *Physica B*, vol. 227, no. 3, pp. 333–335, 1996.

[11] H. Shin, G. M. Yeric, A. F. Tasch, and C. M. Maziar, "Physically-Based Models For Effective Mobility And Local-Field Mobility Of Electrons In MOS Inversion Layers", *Solid State Electonics*, vol. 34, no. 6, pp. 545–552, 1991.

[12] Ronald van Langevelde, and Francois M. Klaasen, "Effect of Gate-Field Dependent Mobility Degradation on Distortion Analysis in MOSFET's", *IEEE Transactions on Electron Devices*, vol. 44, no. 11, pp. 2044–2052, 1997.

[13] K. Y. Lim, and X. Zhou, "A physically-based semi-empirical effective mobility model for MOSFET compact I-V modeling", *Solid State Electonics*, vol. 45, no. 6, pp. 193–197, 2001.

[14] R. Huang, X. Zhang, R. Han, and Y. Wang, "A two-dimensional physically-based current model for deep-submicrometer SOI dynamic threshold-voltage MOSFET", *Solid State Electonics*, vol. 47, no. 8, pp. 1275–1282, 2003.

[15] M. Miura-Mattausch, H. Ueno, H. J. Mattausch, K. Morikawa, S. Itoh, A. Kobayashi, and H. Masuda, "100nm-MOSFET Model for Circuit Simulation: Challenges and Solution", *IEICE Transactions on Electronics*, vol. E86-C, no. 6, pp. 1009–1021, Jun. 2003.

[16] Shin-ichi Takagi, Akira Toriumi, Masao Iwase, and Hiroyuki Tango, "On the Universality of Inversion Layer Mobility in Si MOSFET's: Part II - Effects of Surface Orientation", *IEEE Transactions on Electron Devices*, vol. 41, no. 12, pp. 2363–2368, 1994.

[17] Thomas J. Krutsick, and Marvin H. White, "Consideration of Doping Profiles in MOSFET Mobility Modeling", *IEEE Transactions on Electron Devices*, vol. 35, no. 7, pp. 1153–1155, 1988.

[18] Y P Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill Book Company, 1987.

[19] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[20] G Ghibaudo, "New method for the extraction of MOSFET parameters", *Electronics Letters*, vol. 24, no. 9, pp. 543–545, Apr. 1988.

[21] Man-Chun Hu, and Sheng-Lyang Jang, "An Analytical Fully-Depleted SOI MOSFET Model Considering the Effects of Self-Heating and Source/Drain Resistance", *IEEE Transactions on Electron Devices*, vol. 45, no. 1, pp. 134–148, Jan. 1998.

[22] K Joardar, K K Gullapalli, C C McAndrew, M E Burnham, and A Wild, "An improved MOSFET model for circuit simulation", *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 797–801, Apr. 1998.

[23] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CA94720, *BSIM4.2.0 MOSFET Model - User's Manual*, 2001.

[24] C- L. Huang, and N. D. Arora, "Characterization and modeling of the n- and p-channel MOSFETs inversion-layer mobility in the range 25-125^o", *Solid State Electonics*, vol. 37, no. 1, pp. 97–103, 1994.

[25] Yutao Ma, Litian Liu, Zhiping Yu, and Zhijian Li, "Analytical Charge-Control and I-V Model for Submicrometer and Deep-Submicrometer MOSFETs Fully Comprising Quantum Mechanical Effects", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 4, pp. 495–502, Apr. 2001.

[26] O. Faynot, T. Poiroux, and J. L. Pelloie, "Compact analytical modeling of SOI partially depleted MOSFETs with LETISOI", *Solid State Electonics*, vol. 45, no. 4, pp. 599–605, 2001.

[27] R van Langvelde, *A Compact MOSFET Model for Distortion Analysis in Analog Circuit Design*, University Press Facilities, Eindhoven, 1998.

[28] M. Miura-Mattausch, H. Ueno, M. Tanaka, H. J. Mattausch, S. Kumashiro, T. Yamaguchi, K. Yamashita, and N. Nakayama, "HiSIM: A MOSFET Model for Circuit Simulation Connecting Circuit Performance with Technology", *IEEE International Electron Devices Meeting*, in *IEDM Technical Digest*, 2002, pp. 109–112.

[29] G. Gildenblat, H Wang, T-L Chen, X. Gu, and X Cai, "SP: An Advanced Surface-Potential-Based Compact MOSFET Model", *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1394–1406, Sep. 2004.

[30] Steven A. Schwarz, and Stephen E. Russek, "Semi-Empirical Equation for Electron Velocity in Silicon: Part II-MOS Inversion Layer", *IEEE Transactions on Electron Devices*, vol. 30, no. 12, pp. 1634–1639, 1983.

[31] K. Remashan, N. A. Wong, K. Chan, S. P. Sim, and C. Y. Yang , "Modeling inversion-layer carrier mobilities in all regions of MOSFET operation", *Solid State Electonics*, vol. 46, no. 1, pp. 153–156, 2002.

[32] Mohamed N. Darwish, Janet L. Lentz, Mark R. Pinto, Peter M. Zeitzoff, Thomas J. Krutsick, and Hong Ha Vuong, "An Improved Electron and Hole Mobility Model for General Purpose Device Simulation", *IEEE Transactions on Electron Devices*, vol. 44, no. 9, pp. 1529–1538, 1997.

[33] M. Shirahata, H. Kusano, N. Kotani, S. Kusanoki, and Y. Akasaka, "A mobility Model Including the Screening Effect in MOS Inversion Layer", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 11, no. 9, pp. 1114–1119, Sep. 1992.

[34] G. L. Chindalore, J. B. McKeon, S. Mudanai, S. A. Hareland, W.-K. Shih, C. Wang, A. F. Tasch, and C. M. Maziar, "An Improved Technique and Experimental Results for the Extraction of Electron and Hole Mobilities in MOS Accumulation Layers", *IEEE Transactions on Electron Devices*, vol. 45, no. 2, pp. 502–511, 1998.

[35] M. Kondo, and H. Tanimoto, "An Accurate Coulomb Mobility Model for MOS Inversion Layer and Its Application to NO-Oxynitride Device", *IEEE Transactions on Electron Devices*, vol. 48, no. 2, pp. 502–511, Feb. 2001.

[36] S. Matsumoto, K. Hisamitsu, M. Tanaka, H. Ueno, M. Miura-Mattausch, H. J. Mattausch, S. Kumashiro, T. Yamaguchi, S. Odanaka, and N. Nakayama, "Validity of mobility universality for scaled metal-oxide-semiconductor field-effect transistors down to 100nm gate length", *Journal of Applied Physics*, vol. 92, no. 9, pp. 5228–5232, 2002.

[37] T. Ishihara, and N. Sano, "Coulomb and Phonon Scattering Processes in Metal-Oxide-Semiconductor Inversion Layers: Beyond Mattheissen's Rule", *Japanese Journal of Applied Physics*, vol. 44, no. 4A, pp. 1628–1686, 2005.

Chapter 4

High Field Effects 2: Lateral Field Model

4.1 Introduction

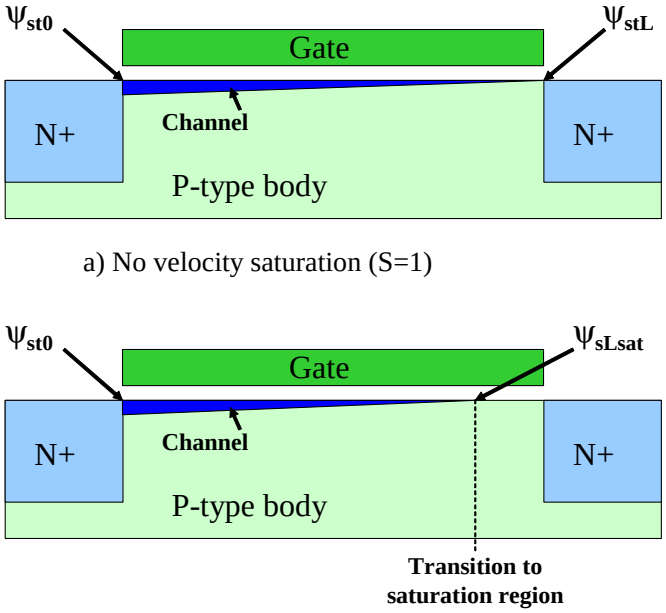

Figure 4.1: *Representation of a MOSFET a) in the linear regime, with the inversion channel extending all the way to the drain terminal, and b) in saturation, with the channel terminating at the transition to the saturation region.*

In the previous chapter, we looked at the issues relevant to the development of a vertical field mobility model, one that could be used to provide an accurate description of the mobility dependence on the vertical component of the electrical field, having first made the assumption that the lateral component is small in comparison. If such an assumption were generally valid then the construction of MOSFET compact models would be greatly simplified. However, this is only true for low values of drain-source voltage; beyond a

cerain point, quickly reached in modern sub-micron devices, the lateral electric field becomes too large for its influence to be ignored.

Figure 4.1 shows how a high lateral field affects device operation. In Figure 4.1 a), we depict a NMOS device, biased in strong inversion. The channel extends all the way to the drain, and the drain surface potential is simply ψ_{stL}. In Figure 4.1 b), we show the same device, but with the drain voltage increased to the point where velocity saturation of the carriers has occured. The channel now terminates before reaching the drain; the surface potential at this point is ψ_{sLsat}, the drain saturation surface potential. This concept of a saturation surface potential was first published in [1], and was used in the original STAG2 model [2].

In effect, ψ_{sLsat} is the maximum value that ψ_{sL} can attain in the model. Assuming that $V_{DS} \geq 0$, the surface potential at the drain end cannot be less than that at the source end, being $\psi_{s0} \approx \psi_{st0}$. Also, if no velocity saturation effect were to be included, then the drain saturation potential would be attained when the channel charge density q_c vanishes. Using these criteria and the linearised channel charge expression in Equation (2.19), the upper bound of the saturation potential is found to be $\frac{VGT}{\alpha}$. Therefore, the saturation potential including high field mobility effects is defined in STAG2 as [2]

$$\psi_{sLsat} = \psi_{st0} + \frac{\Psi}{S} \tag{4.1}$$

where $1 < S \leq \infty$, and $\Psi = (V_{GT}/\alpha) - \psi_{st0}$. In the limit of no velocity saturation ($S = 1$), $\psi_{sLsat} = (V_{GT}/\alpha)$, whilst in the limit of excessive velocity saturation ($S \to \infty$), ψ_{sLsat} tends asymptotically to ψ_{st0}. Thus, it is ensured that ψ_{sLsat} is limited to a physical range of values.

4.2 Modelling Velocity Saturation

The STAG model uses an expression for the channel current I_{CH} of the form:

$$I_{CH} = \frac{WC_{of}}{L}\mu_{eff}\left[f\left(\psi_{sL}\right) - f\left(\psi_{s0}\right)\right] \tag{4.2}$$

where μ_{eff} is the total effective mobility, and $f(\psi_s)$ is given by (2.30). Most compact models, including STAG, take the same basic approach to modelling μ_{eff}. The following empirical relation is commonly used [3–6]

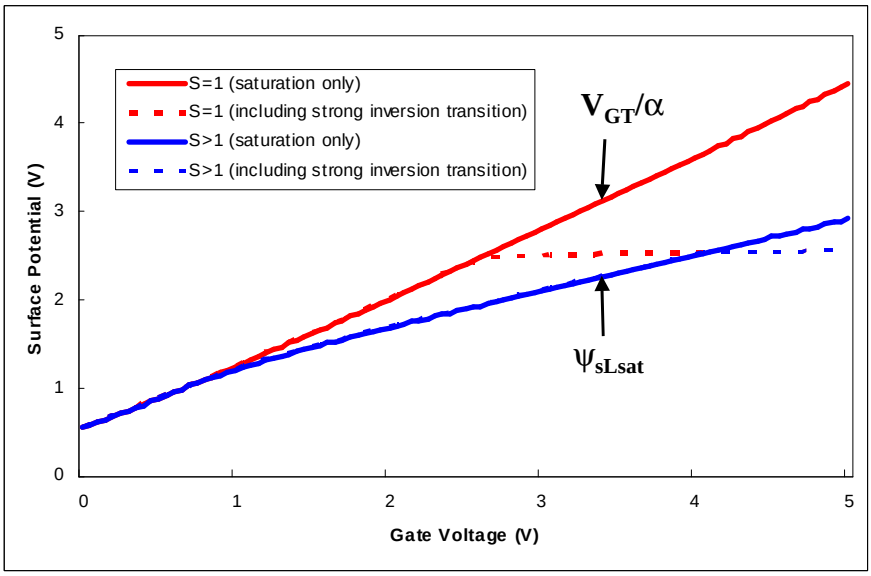

Figure 4.2: *Plots of surface potential against gate voltage, with (red line) and without (blue line) the effects of velocity saturation. Note that the difference is only seen outside strong inversion - at the appropriate value of gate voltage, both curves transition into the dotted lines.*

$$\mu_{eff} = \frac{\mu_{xeff}}{\left[1 + \left(\frac{E_{yeff}}{E_c}\right)^{\beta}\right]^{\frac{1}{\beta}}} \tag{4.3}$$

where E_{yeff} is the average lateral field experienced by carriers in the channel, E_c is the critical lateral field corresponding to the onset of velocity saturation, and β is normally a fitting parameter. E_c is generally defined as

$$E_c = \frac{v_{sat}}{\mu_{xeff}} \tag{4.4}$$

where v_{sat} is the carrier saturation velocity. It will be recalled that μ_{xeff} is the vertical field limited mobility, an expression for which has already been derived for STAG in Equation (3.25). Substituting Equation (4.4) into Equation (4.5) gives us

$$\mu_{eff} = \frac{\mu_{xeff}}{\left[1 + \left(\frac{\mu_{xeff} E_{yeff}}{v_{sat}}\right)^{\beta}\right]^{\frac{1}{\beta}}} \tag{4.5}$$

There has over the years been a certain amount of controversy over the 'true' values of v_{sat} and β. There has been a kind of general consensus that β should be set to 1 for PMOS and 2 for NMOS [5]. The established literature value for v_{sat} for electrons is 1×10^7 cms^{-1}. Compact models, on the other hand, sometimes use β=1 for both cases, on the

grounds that it makes the mathematics more tractable, with v_{sat} usually being employed as a fitting parameter. Various groups have made attempts to resolve this issue. Taking the case of electron mobility, around which the majority of previous work has been centred, Assaderaghi et al. found a value for v_{sat} of around 8.5×10^6 cms^{-1}, and for β of 1.4 [3]. Roldan et al. concluded that the optimum value for v_{sat} was $1.1 \times 10^7 cms^{-1}$, and that β was not a constant but was itself a function of E_{yeff} [4]. A more recent, and possibly more rigorous, study by Hoyniak et al. found that a unity value for β gave the best experimental fit for both NMOS and PMOS devices, but that v_{sat} was somewhat lower than most other literature values, at 4×10^6cms^{-1} [5]. A distinguishing feature of this work is that the range of E_{yeff} over which the fitting was performed exceeds that of any previous study by a factor of 5-10. Such a large range is indeed required if the whole operating range of a deep-submicron device is to be accounted for; furthermore the results seem to indicate that neglecting the upper part of the E_{yeff} range may very well lead to an overestimation of both v_{sat} and β.

The source of all this confusion has been put down to numerous reasons, including differences in extraction technique, different definitions of v_{sat}, and different technologies being examined. This last one is particularly problematic, since other short-channel effects can easily affect the results unless they are properly accounted for in the model, and these effects can vary considerably between different technologies. One possible source of error which has received little discussion is that the velocity saturation model and its parameters is closely tied to the vertical field model. As was discussed in Chapter 3, vertical field models have evolved considerably over the past 5-10 years, and quite a number of variations are now in use. This lack of consistency, coupled with the general failure to provide details of the vertical field expressions, makes rigorous evaluation of the published literature very difficult.

4.3 STAG Lateral Field Model

4.3.1 Series Resistance

STAG2 already provides the facility to model source and drain series resistance externally. However, these external resistances introduce two internal nodes into the devices MOSFET sub-circuit model. This is acceptable when simulating small circuits or simple devices, but with larger circuits we might expect simulation time to increase significantly. Furthermore, apart from the algorithms introduced to scale the series resistances automatically with device width (discussed in Chapter 7), these nodal resistances are constant, quite independent of terminal bias.

With the emergence of drain engineered devices such as Lightly Doped Drain (LDD) MOSFETs, the series resistance can have a strong dependence on the drain and particularly the gate bias. In order to model this dependence, and also to eliminate superfluous inter-

nal nodes, a new internal model was introduced [7], with separate equations for R_S and R_D.

$$R_S = R_{Sint}\left(1 + \frac{a_{R1}}{a_{R2} + V_{GfB} - V_{FB}}\right) \tag{4.6}$$

$$R_D = R_{Dint}\left(1 + \frac{a_{R1}}{a_{R2} + V_{GfB} - V_{FB}}\right) \tag{4.7}$$

where R_{Sint}, R_{Dint}, a_{R1}, and a_{R2} are fitting parameters. R_{Sint} and R_{Dint} are inversely proportional to W. Adding the source and drain contributions gives the total series resistance f_R (note that because f_R appears in the denominator of some equations later in this chapter, R_{Sint} and R_{Dint} are set by default to a small, non-zero value to avoid numerical problems).

$$f_R = R_S + R_D \tag{4.8}$$

Finally, again following the approach in [7], we include the series resistance contribution as a term in the denominator of the mobility expression. Equation (3.25) is therefore modified to become

$$\mu_{xeff} = \frac{\mu_0}{G_v + f_R} \tag{4.9}$$

4.3.2 Velocity Overshoot

So far, our discussion of carrier mobilities have assumed that the device is operating under the standard drift-diffusion model for carriers. In other words, the carriers are in thermal equilibrium with the silicon lattice, exchanging energy with it as they undergo scattering events. The maintenance of this thermal equilibrium is what causes the drift velocity of the carriers to be restricted to a particular maximum value, which we already know as the saturation velocity.

In deep sub-micron devices, however, the electric fields become sufficiently strong that the drift-diffusion model no longer fully applies. The carriers receive enough energy from the lateral electric field that they remain out of thermal equilbrium with the lattice. The carriers are now said to be undergoing *ballistic transport*. The result is an increase in the device transconductance, which exceeds the theoretical maximum value applicable to a drift-diffusion only model. The effect is termed *velocity overshoot*.

In order to model this effect, we start with the treatment given in [8]. The increase in the local field-defined velocity in the drift-diffusion model, v_D, is given by

$$v = v_D \left(1 + \frac{k_B T \mu_{eff0}}{q v_{sat} E_{yeff}} \frac{dE_{yeff}}{dy}\right) \tag{4.10}$$

At high values of E_{yeff}, v_D approaches v_{sat}. We can therefore apply Equation (4.10) to v_{sat} to obtain an new effective value which we will call v_{sat_vo}

$$v_{sat\_vo} = v_{sat} \left(1 + \frac{k_B T \mu_{xeff}}{q v_{sat} E_{yeff}} \frac{dE_{yeff}}{dy}\right) \tag{4.11}$$

We now need to find expressions for E_{yeff} and $\frac{dE_{yeff}}{dy}$. For this, we use the following standard expressions [9]

$$E_{yeff} = \frac{(\psi_{sL} - \psi_{s0})}{L} \tag{4.12}$$

$$\frac{dE_{yeff}}{dy} = \frac{a_{vo}\,(\psi_{sL} - \psi_{s0})}{L^2} \tag{4.13}$$

where a_{vo} is a constant of propportionality. Substituting Equations (4.12) and (4.13) into Equation (4.11) gives

$$v_{sat\_vo} = v_{sat} + a_{vo} \frac{k_B T \mu_{xeff}}{qL} \tag{4.14}$$

We now lump the various constants into a single model parameter λ_{vo}

$$v_{sat\_vo} = v_{sat} + \lambda_{vo} \frac{\mu_{xeff} T}{L} \tag{4.15}$$

Finally we need to simply the expression by replacing μ_{xeff} with μ_0

$$v_{sat\_vo} = v_{sat} + \lambda_{vo} \frac{\mu_0 T}{L} \tag{4.16}$$

4.3.3 Final Expression

We need to obtain an expression for μ_{eff} in terms of the surface potential. Most of the work has already been done; all we have to do is to substitute μ_{xeff} and E_{yeff}, from Equations (4.9) and (4.12) respectively, into Equation (4.5), and replace v_{sat} with v_{sat_vo} from Equation (4.16). Simplifying then yields the following expression

$$\mu_{eff} = \frac{\mu_0}{G_v\sqrt{\left[1+\left(\frac{\mu_0(\psi_{sL}-\psi_{s0})}{G_v(v_{sat}L+\lambda_{vo}\mu_0 T)}\right)^2\right]}+f_R} \tag{4.17}$$

4.4 Calculation of Saturation Surface Potential

The standard approach to solving ψ_{sLsat} is to equate it to the drain surface potential at which the channel current turns over (reaches saturation) [10]. Therefore, $\psi_{sL} = \psi_{sLsat}$ when the following condition is met:

$$\frac{\partial I_{CH}}{\partial \psi_{sL}} = 0 \tag{4.18}$$

If ψ_{sLsat} is over-estimated, it can result in a non-physical rollover of the channel current [1]. Figure 4.3 shows the effect of this on the device characteristics. Problems can occur even if the over-estimation is just a few mV. So, while it is therefore important to ensure that ψ_{sLsat} is calculated as accurately as possible, we must give the highest priority to ensuring that our value does not even slightly over-estimate the exact solution of Equation (4.18).

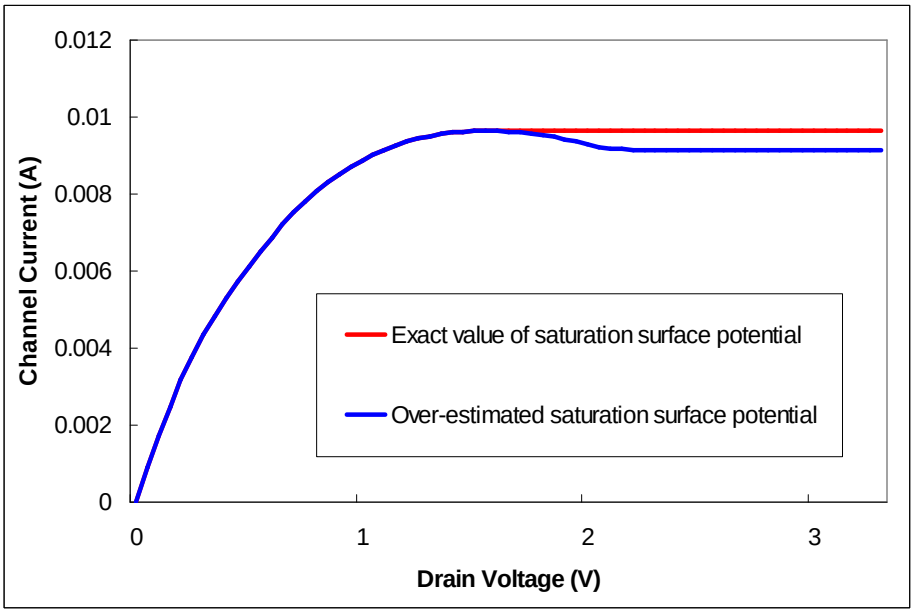

Figure 4.3: *Plot of channel current against gate voltage, showing the non-physical behaviour that occurs if the saturation surface potential is overestimated.*

It is interesting to note that little coverage is given to the determination of ψ_{sLsat} in the literature. Older empirical models did not have to worry about this, being piece-

wise models which often contained discontinuities at the boundaries between operating regimes. Early surface potential models were often formulated using simple expressions, so that (4.18) yielded a quadratic expression for ψ_{sLsat} [1,11]. But newer models employing more complex mobility relations cannot go down this route without sacrificing some of their accuracy. One possibility is to approximate with a series expansion, but there is not necessarily any guarantee that the resulting expression will avoid the roll-over problem.

Another approach is to retain the effect of vertical field degradation when calculating I_{CH}, but neglect it when calculating ψ_{sLsat} [7]. This has the benefit of avoiding roll-over; because vertical mobility degradation acts to delay the onset of saturation, and thus raises the drain surface potential at which it does occur, removing it from the calculation ensures that the approximate solution for ψ_{sLsat} will always be lower than the exact solution of Equation (4.18). Because the G_v term in Equation (4.17) is being set to 1, the only functional dependence on ψ_{sL} comes from the velocity saturation term (the second term in the denominator). The result of the differentiation, after making suitable approximations, is therefore an easily solvable quadratic equation.

Without further ado, let us proceed with the treatment for obtaining ψ_{sLsat} in the STAG3 model. We will make the popular assumption that β should be set to 1 for PMOS and 2 for NMOS in Equation (4.5), and we will consider the more complex case of $\beta = 2$.

Attempting to directly substitute Equations (4.2) and (4.17) into Equations (4.18) leads to a very complicated expression. This is because G_v in (4.17) has a complex functional dependence on the surface potential (due to there being three contributing terms: G_{sr}, G_{ph}, and G_{cou}). We would like to simplify the mathematics, but we would prefer not to completely ignore the effects of vertical field degradation by setting G_v to 1 as in [7]. A compromise is to replace G_v with G_{v0}, the mobility reduction factor at the source end of the channel. Since we are now using ψ_{s0}, the surface potential at the source, instead of the average surface potential along the channel, G_{v0} will be a constant, and will not vary with ψ_{sL}. In other words

$$\frac{\partial G_{v0}}{\partial \psi_{sL}} = 0 \tag{4.19}$$

Thus we can simplify the mathematics considerably, whilst still getting some kind of estimate for the effect of the vertical field in delaying the onset of velocity saturation. Since the magnitude of G_{v0} is less than G_v, we still ensure that our approximate ψ_{sLsat} will be lower than the exact solution of Equation (4.18). Hence, we are guaranteed to avoid roll-over, though our error will be smaller than [7]; clearly, using $G_v = G_{v0}$ is more accurate than simply setting $G_v = 1$.

If we set $G_v = G_{v0}$ in Equation (4.17), we can then re-write Equation (4.2) in the form

$$I_{CH} = \frac{WC_{of}}{L} \frac{\mu_0}{yG_{v0} + f_R} \left[f\left(\psi_{sL}\right) - f\left(\psi_{s0}\right) \right] \tag{4.20}$$

where

$$y_{sat} = \sqrt{1 + \left(\frac{\mu_0 \left(\psi_{sL} - \psi_{s0}\right)}{G_{v0} \left(v_{sat} L + \lambda_{vo} \mu_0 T\right)} \right)^2} \tag{4.21}$$

If we now apply Equation (4.18) to Equation (4.20), we obtain the following

$$\frac{dI_{CH}}{d\psi_{sL}} = \frac{WC_{of}}{L} \left(\mu_{eff} \frac{df\left(\psi_{sL}\right)}{d\psi_{sL}} + \frac{d\mu_{eff}}{d\psi_{sL}} \left[f\left(\psi_{sL}\right) - f\left(\psi_{s0}\right) \right] \right) = 0 \tag{4.22}$$

Of course, we need to derive expressions for $\frac{df(\psi_{sL})}{d\psi_{sL}}$ and $\frac{d\mu_{eff}}{d\psi_{sL}}$. From Equation (2.30), it can easily be seen that

$$\frac{df\left(\psi_{sL}\right)}{d\psi_{sL}} = V_{GBT} - \alpha \psi_{sL} \tag{4.23}$$

Similarly, we find that

$$\frac{d\mu_{eff}}{d\psi_{sL}} = \frac{-\mu_0}{\left(y_{sat} G_{v0} + f_R\right)^2} \left(G_{v0} \frac{dy}{d\psi_{sL}} + \frac{dG_{v0}}{d\psi_{sL}} y_{sat} \right) \tag{4.24}$$

where from Equation (4.21)

$$\frac{dy_{sat}}{d\psi_{sL}} = \frac{\mu_0^2 \left(\psi_{sL} - \psi_{s0}\right)}{G_{v0}^3 \left(v_{sat} L + \lambda_{vo} \mu_0 T\right)^2} \left[G_{v0}^2 - \frac{dG_{v0}}{d\psi_{sL}} \left(\psi_{sL} - \psi_{s0}\right) \right]$$

$$\sqrt{\frac{G_{v0}^2 \left(v_{sat} L + \lambda_{vo} \mu_0 T\right)^2}{G_{v0}^2 \left(v_{sat} L + \lambda_{vo} \mu_0 T\right)^2 + \mu_0^2 \left(\psi_{sL} - \psi_{s0}\right)^2}} \tag{4.25}$$

If we substitute Equation (4.23) and Equation (4.24) into Equation (4.22), we get

$$\frac{\mu_0}{y_{sat} G_{v0} + f_R} \left(V_{GBT} - \alpha \psi_{sL} \right)$$

$$- \frac{\mu_0}{\left(y_{sat} G_{v0} + f_R\right)^2} \left[G_{v0} \frac{dy_{sat}}{d\psi_{sL}} + \frac{dG_{v0}}{d\psi_{sL}} y_{sat} \right] \left[f\left(\psi_{sL}\right) - f\left(\psi_{s0}\right) \right] = 0 \tag{4.26}$$

Now multiply both sides by $\frac{y}{\mu_0} \left(y G_{v0} + f_R\right)^2$ and apply Equation (4.19)

$$\left(y_{sat}^2 G_{v0} + y_{sat} f_R\right)\left(V_{GBT} - \alpha\psi_{sL}\right) - y_{sat} G_{v0} \frac{dy}{d\psi_{sL}}\left[f\left(\psi_{sL}\right) - f\left(\psi_{s0}\right)\right] = 0 \tag{4.27}$$

When Equation (4.27) is expanded out, we find that we now have only a single term containing a fractional power of ψ_{sL}. We therefore take a Taylor expansion of $y_{sat} f_R$, and then solve the resulting polynomial. Since $\psi_{sL} = \psi_{sLsat}$ at roll-over, we introduce it now in place of ψ_{sL}.

$$\psi_{sLsat}^3 + a_1 \psi_{sLsat}^2 + a_2 \psi_{sLsat} + a_3 = 0 \tag{4.28}$$

where

$$a_1 = -3\psi_{s0} + \frac{2}{\mu_0^2}\left[G_{v0}\left(v_{sat} L + \lambda_{vo}\mu_0 T\right)^2\right] z_{sat} f_R \tag{4.29}$$

$$a_2 = 3\psi_{s0}^2 + \frac{2}{\mu_0^2}\left[G_{v0}\left(v_{sat} L + \lambda_{vo}\mu_0 T\right)^2\right]\left[G_{v0} + y_{sat} f_R - z_{sat} f_R\left(\psi_{ss} + \frac{V_{GT}}{\alpha}\right)\right] \tag{4.30}$$

$$a_3 = -\psi_{s0}^3 - \frac{2}{\mu_0^2}\left[G_{v0}\left(v_{sat} L + \lambda_{vo}\mu_0 T\right)^2\right]\frac{V_{GT}}{\alpha}\left[G_{v0} + y_{sat} f_R - z_{sat} f_R \psi_{ss}\right] \tag{4.31}$$

and

$$z_{sat} = \frac{\mu_0^2\left(\psi_{sL} - \psi_{s0}\right)}{y_{sat} G_{v0}^2\left(v_{sat} L + \lambda_{vo}\mu_0 T\right)^2} \tag{4.32}$$

The problem with deriving ψ_{sLsat} in this way is that it does not ensure that ψ_{sLsat} remains in the physical range discussed in Section 4.1 ($\psi_{st0} \leq \psi_{sLsat} \leq \frac{VGT}{\alpha}$). Attempts to use this formulation in STAG3 resulted in non-physical results and numerical evaulation errors. In particular, problems were encountered when the body potential was greater than the source and drain potentials ($V_{SB} \leq V_{DB} < 0$).

In order to resolve this problem, we use Equation (4.1) to confine ψ_{sLsat} to a physical range of values, just as was done in STAG2. We substitute Equation (4.1) into Equation (4.28), multiply by S^3, and simplify to obtain the following cubic equation in S

$$S^3 + a_1 S^2 + a_2 S + a_3 = 0 \tag{4.33}$$

where

$$a_1 = A_{sat} - 1 \tag{4.34}$$

$$a_2 = -A_{sat} \tag{4.35}$$

$$a_3 = -A_{sat}\left(\frac{G_{vsat}\Psi}{2G_{v0}f_R C_1}\right) \tag{4.36}$$

and

$$A_{sat} = \frac{f_R C_1 \Psi}{G_{v0} + f_R\left(C_{0+}C_1\left[\psi_{ss} - \psi_{st0}\right]\right)} \tag{4.37}$$

$$C_0 = \sqrt{1 + \frac{G_{vsat}(\psi_{ss} - \psi_{st0})^2}{G_{v0}}} \tag{4.38}$$

$$C_1 = \frac{G_{vsat}\left(\psi_{ss} - \psi_{st0}\right)}{C_0 {G_{v0}}^2} \tag{4.39}$$

Using Equation (4.33)-(4.39) allows us to compute S in a robust manner. Converting to ψ_{sLsat} is simply a matter of using S in Equation (4.1). We shall return to the inclusion of ψ_{sLsat} in our surface potential model at the end of Chapter 5, once our high field treatment has been concluded.

References

[1] A R Boothroyd, S W Tarasewiez, and C Slaby, "MISNAN - a physically based continuous MOSFET model for CAD applications", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 10, no. 12, pp. 1512–1529, Dec. 1991.

[2] M S L Lee, B M Tenbroek, W Redman-White, J Benson, and M Uren, "A physically based compact model of partially depleted SOI MOSFETs for analogue circuit simulation", *IEEE Journal of Solid-State Circuits*, vol. 36, no. 1, pp. 110–121, Jan. 2001.

[3] Fariborz Assaderaghi, Dennis Sinitski, Jeffrey Bokor, Ping K. Ko, Henry Gaw, and Chenming Hu, "High-Field Transport of Inversion-Layer Electrons and Holes Including Velocity Overshoot", *IEEE Transactions on Electron Devices*, vol. 44, no. 4, pp. 664–671, 1997.

[4] J. B. Roldan, F. Gamiz, J A. Lopez-Villanueva, J. E. Carceller, and P. Cartujo, "The dependence of the electron mobility on the longitudinal electric field in MOSFETs", *Semiconductor Science and Technology*, vol. 12, no. 3, pp. 321–330, 1997.

[5] Dennis Hoyniak, Edward Nowak, and Richard L. Anderson, "Channel electron mobility dependence on lateral electric field in field-effect transistors", *Journal of Applied Physics*, vol. 87, no. 2, pp. 876–881, 1997.

[6] J. A. Lopez-Villanueva, F. Gamiz, J. B. Roldan, Y. Ghailan, J. E. Carceller, and P. Cartujo, "Study of the effects of a stepped doping profile in short-channel MOSFET's", *IEEE Transactions on Electron Devices*, vol. 44, no. 9, pp. 1425–1431, 1997.

[7] R van Langvelde, *A Compact MOSFET Model for Distortion Analysis in Analog Circuit Design*, University Press Facilities, Eindhoven, 1998.

[8] Lixin Ge, Jerry G. Fossum, and Bin Yiu, "Physical Compact modeling and Analysis of Velocity Overshoot in Extremely Scaled CMOS Devices and Circuits", *IEEE Transactions on Electron Devices*, vol. 48, no. 9, pp. 2074–2080, 2001.

[9] J. B. Roldan, F. Gamiz, J. A. Lopez-Villanueva, and J. E. Carceller, "Modeling effects of Electron-Velocity Overshoot in a MOSFET", *IEEE Transactions on Electron Devices*, vol. 44, no. 5, pp. 841–846, May 1997

[10] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[11] R Howes, W Redman-White, K G Nichols, P J Mole, M J Robinson, and S Bird, "An SOS MOSFET model based on calculation of the surface potential", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 6, pp. 494–506, Apr. 1994.

Chapter 5

High Field Effects 3: Quantum Mechanical Model

5.1 Introduction

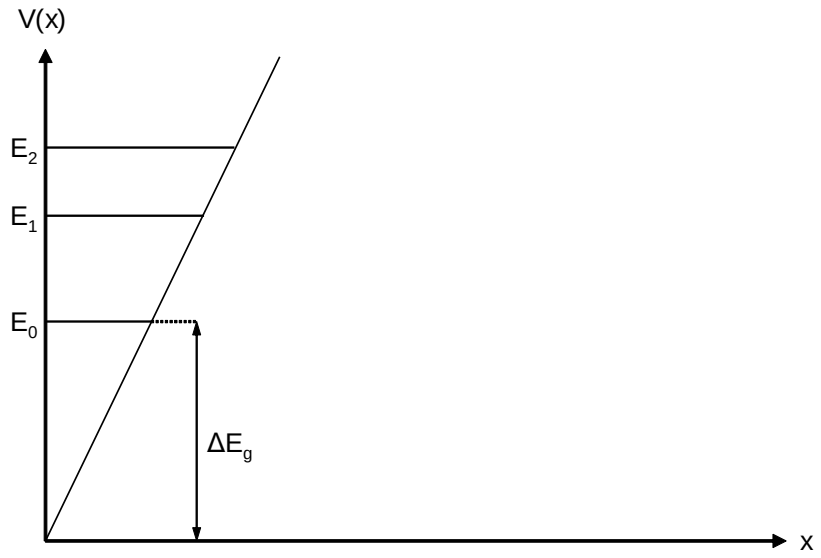

Figure 5.1: *Representation of a quantum well formed by strong conduction band-bending close to the silicon-oxide interface. The lowest three quantised energy levels are also shown.*

As devices have been scaled down into the deep-submicron regime, it has been necessary to increase the doping of the silicon layer in order to control threshold voltage and to suppress short-channel effects. This has in turn led to an increase in the vertical electric field being applied across the inversion layer. With increased doping levels, it becomes possible for the gate field to be sufficiently strong - even around the inversion threshold - to induce significant band-bending in the conduction band (or valence band in the case of

holes) close to the silicon-oxide interface. When the width of the confining potential becomes comparable with the thermal de Broglie wavelength of the carriers in the inversion channel, those carriers find themselves subject to quantum mechanical confinement in the direction perpendicular to the interface [1]. In effect a quantum well is created between the interface and the conduction band. This situation is illustrated in Figure 5.1.

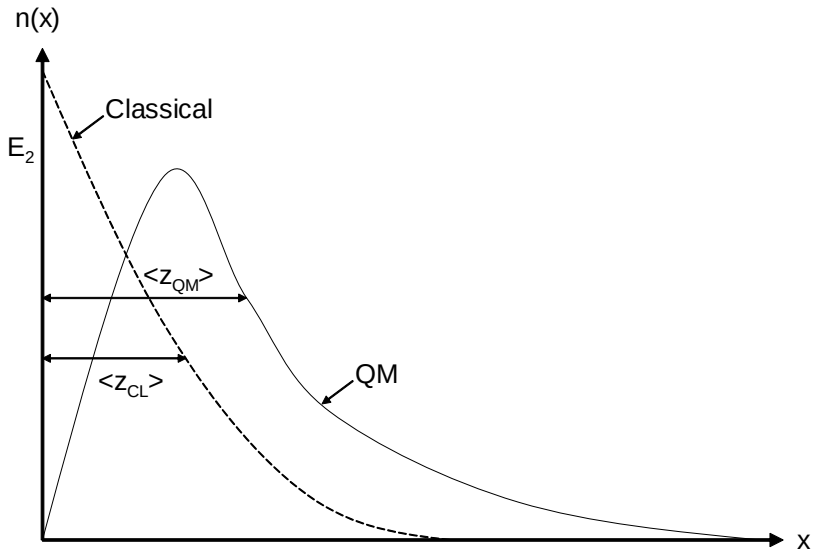

Figure 5.2: *Comparison of the classical and quantum mechanical distribution of inversion layer electrons as a function of depth into the silicon. z_{CL} and z_{QM} give the average distance of electrons from the silicon-oxide interface in each case.*

This quantisation of the inversion layer carriers leads to two important effects, both of which arise from basic quantum well theory. The first is that the perpendicular component of the carrier energies can no longer take on a continuum of values, but instead are restricted to certain quantised states. The lowest of these energy states, the ground state, does not lie on the conduction band, but is displaced by some energy value ΔE_g. Since carriers now require more energy to enter the conduction band, and hence contribute to the inversion charge, we see a reduction in the inversion charge density for a given surface potential [2]. This can also be viewed as an effective increase in the threshold voltage [3–5].

The second effect relates to the distribution of carriers within the inversion layer. It is usual for compact models to assume that the charge sheet model [6] is valid. In other words, the inversion charge consists of an infinitessimally thin conducting layer located at the silicon-oxide interface. This is consistent with the classical nature of the assumptions underlying these models. In a confined system, it is no longer valid to assume that the entire carrier population is located at the interface. Instead, it is distributed through the

quantum well, and the centroid is displaced away from the interface by some distance Δz_I [2]. This quantity was introduced in Equation (3.2), and results in an effective increase in the gate oxide width [3, 4].

Figure 5.2 depicts the situation for electron distribution in the semi-classical and quantum mechanical regimes. In situations where quantum effect are ignored, it is assumed that $\Delta z_I = \Delta z_{CL}$. Furthermore, it is often common practice to assume that $\Delta z_{CL} = 0$ - in other words that the entire carrier population is located in an infinitesimally thin charge layer located at the silicon-oxide interface. This assumption is acceptable for situations where $t_{of} \gg \Delta z_{QM}$, since the effective increase in the oxide thickness is small compared to the real value. However, in modern MOSFET devices, t_{of} is typically only a few nm, while Δz_{QM} has a field dependent value which is of the order of $\sim$1nm. Thus, it now becomes preferable to set $\Delta z_I = \Delta z_{QM}$.

Another difference of the quantum confinement model compared to semi-classical one is depicted in Figure 5.3. Here we plot the density of states (DOS) for the electron population in a NMOS inversion layer against the carrier energy (with the x-axis origin taken as E_c, the semi-classical value of the bottom of the conduction band). We see that the classical DOS distribution is a continuous function that begins at E_c. In contrast, the quantum mechanical distribution matches a 2-D step-like DOS profile, corresponding to a number of discrete subbands, each of of which has constant electron occupancy over the energy range pertaining to that subband [2]. In the literature related to quantum confinement of the inversion layer, many treatments begin with the assumption that the inversion layer is operating in the *electrical quantum limit*, when all the inversion carriers occupy the lowest subband. This occurs in the limit of high electric fields and low operating temperatures, such that the thermal carrier energy k_BT is small compared to the separation between the subbands. In real devices this assumption is generally not valid, resulting in errors when predicting deviations from the semi-classical case.

Note that Figure 5.3 is only intended to give a idea of the qualitative differences between the 3-D and 2-D DOS profiles. For more precise profiles, together with a more detailed treatment of the influence of how changes in the DOS structure influence device behaviour, references such as [7] should be consulted.

5.2 Quantum Theory of MOSFET Inversion Layers

Much of the work in this area uses results from a few early papers which developed approximate expressions to determine the quantized energy states in a MOS inversion layer [1, 2]. A lengthy but comprehensive review paper gives extensive background on these and other early studies [8]. Due to the complexity of many real quantum systems, it is usual for self-consistent computational calculations to be employed. Where a high degree of accuracy is needed, this holds true for silicon inversion layers as well. However,

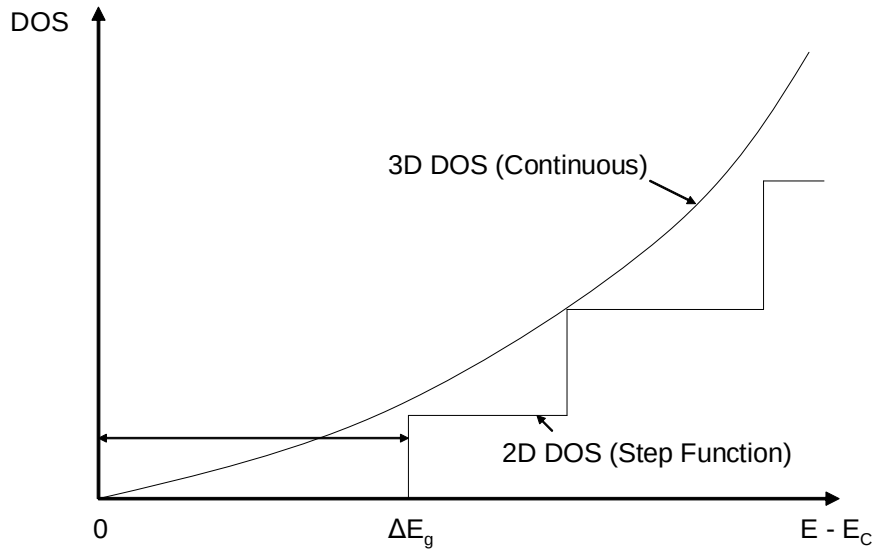

Figure 5.3: *Representation of the density of states (DOS) for 2D and 3D electron distributions*

given some approximations, it is possible to reduce the complexity to the point where analytical expressions can be obtained for important quantities. In one early paper by Stern [2] two methods are summarised which have yielded particularly useful results for subsequent work. These are the triangular potential approximation and the variational method.

5.2.1 Triangular Potential Approximation

Strictly speaking, the band bending created close to the interface follows a parabolic curve. Accurate results are obtained when simultaneously solving the Schrodinger and Poisson equations using a parabolic potential profile [9]. However, numerical iteration is needed to arrive at the solution. The same is true if a logarithmic potential is used instead [10]. A useful simplification is to replace the parabolic potential with a triangular one. This permits an analytical solution, if a second simplifying assumption is also made - that the interface is an infinite barrier. In other words. it is assumed that there is no penetration of the carrier wavefunction into the oxide.

Under these assumptions, the expression for the energy of a particular quantised state is found to be [2]

$$E_{ij} = \left(\frac{\hbar^2}{2m_i}\right)^{1/3} \left[\frac{3}{2}\pi q E_{xeff}\left(j + \frac{3}{4}\right)\right]^{2/3} \tag{5.1}$$

where E_{xeff} is the effective vertical gate field defined in Equation (3.1), E_{ij} is the jth

energy level in the ith valley (referring to the energy bands), and m_i is the effective mass in that valley. Since E_{ij} is expressed relative to the bottom of the conduction band, we can find ΔE_g by setting it equal to the lowest energy state

$$\Delta E_g = E_{10} = \left(\frac{\hbar^2}{2m_1}\right)^{1/3} \left[\frac{9}{8}\pi q E_{xeff}\right]^{2/3} \tag{5.2}$$

While the triangular well approximation is an appealing prospect because it offers closed-form solutions, the question must be asked: how much accuracy is being lost? Let us consider the infinite barrier approximation first. A recent study by Kauser et al [11] has showed that this is a valid assumption, with the associated error being typically no more than a few percent. We can therefore treat this as an acceptable simplification. However, the triangular well approximation itself is more of a problem. Using this simplified profile leads to a significant loss of accuracy in the case where the inversion charge density is comparable with the body charge density [2]. This is unfortunate, since it means that the greatest loss of accuracy occurs in strong inversion, exactly where quantum effects have the most pronounced effect on device behaviour. The treatment is still accurate in depletion and weak inversion, since here the inversion charge density is too small to perturb the potential away from the triangular case. A study by Ma et al [12], in which an analytical triangular well solution was compared with a numerical simulation based on the parabolic model, demonstrated that the shift in surface potential was very similar in both cases. However, the triangular well treatment did show poorer accuracy when calculating the inversion layer charge centroid at high gate overdrives (i.e. well into strong inversion), where the error was as high as a factor of 2 or more.

The work described in [12] raises another point. The triangular well approximation does not actually give a convenient result for the position of the peak inversion charge. In fact, only the initial solution of the Schrodinger equation to obtain the wavefunctions and energy states was done analytically for the triangular case in [12]. Numerical calculation was then used to determine the inversion charge centroid. Obviously this must be considered a major disadvantage from a compact modelling perspective.

5.2.2 Variational Method

Having looked at the triangular potential approximation, let us turn our attention to the second treatment, which is based on the variational method. Once again it is assumed that the interface forms an infinite potential barrier. It is also assumed that all carriers reside in the lowest subband i.e. in the electrical quantum limit when the thermal carrier energy k_BT is small compared to the separation between the subbands. Using this method, ΔE_g is found to be [2]

$$\Delta E_g = \frac{3\hbar^2 b^2}{8m_e} \tag{5.3}$$

where b is the variational parameter and is given by

$$b = \left(\frac{12 m_e q}{\epsilon_{si} \hbar^2}\right)^{1/3} \left[q_b + \frac{11}{32} q_c\right]^{1/3} \tag{5.4}$$

The variational method also provides us with a simple expression for Δz_I, the displacement of the inversion layer charge centroid away from the silicon-oxide interface [2]

$$\Delta z_I = \frac{3}{b} \tag{5.5}$$

These results are valid in strong inversion, but lose accuracy in medium and weak inversion [2]. However, the effects of quantum confinement become less important once we move outside the strong inversion region of operation anyway; since the carrier density falls rapidly outside strong inversion, the vertical electric field is no longer sufficiently large to impose strong confinement on the remaining carriers. From a compact modelling perspective, it is sufficient to achieve an accurate functional dependency in strong inversion, and then employ some empirical expression to ensure a gradual reduction in the contribution from quantum mechanical effects.

5.3 Quantum Mechanical Compact Modelling

The variational approach seems to be favoured by several groups interested in constructing quantum mechanical models suitable for circuit simulation. Pregaldiny et al [13, 14] used Equations (5.3) and (5.4) as the basis for their quantum model of inversion layers. Interestingly, they extended their model to account for accumulation as well, but decided to use Equation (5.2), since the occupation of the energy levels becomes more complicated in this case [15]. Since STAG3 does not account for the accumulation regime, this is outside of the scope of the present work.

Clerc et al also used Equations (5.3) and (5.4) as the basis for their compact model for direct tunnelling from MOSFET inversion layers [16]. They developed separate models for weak and strong inversion, the need for accuracy in both regimes being a consequence of their specific application for quantum tunelling. Again, this is outside the scope of the development of STAG3.

When we look at well-established compact models, it can be seen that they tend to adopt quite simple quantum mechanical expressions. We have seen that there are two main effects arising from carrier confinement - lowering of the inversion charge for a given gate voltage, and an increase in the effective gate oxide thickness. What is interesting is that many compact models will only model one effect or the other, in effect lumping both effects into the same set of equations. This is probably partly a consequence of a frequently referenced study by Van Dort et al [4], in which the two effects were lumped into a single term which was then used to model the threshold voltage shift associated with quantum

confinement. It was later argued by Ma et al [7] that this approach - taking one effect relevant to energy space, and another relevant to real space - and lumping them together, was non-physical, and could lead to significant over-estimation of the threshold voltage shift. This is perhaps too subtle a point to apply to a circuit simulation model, where rigorously derived constants of proportionality are often converted into fitting parameters, but it is something to keep in mind as we briefly review the approaches taken by some of the well-known compact models. We shall include key equations for each model; even if the treatments are too brief to allow proper comparison, they at least allow the reader to get a feel for commonly employed expressions. Where appropriate, we will highlight similarities between these model equations and the equations given earlier in the chapter.

The HiSIM model only models the charge centroid displacement effect, using the assumptions of a triangular well and carriers being restricted to the lowest energy subband. The resulting effective gate oxide thickness is given by [17]

$$t_{oxeff} = t_{ox} + \alpha \left(Q_b + \frac{11}{32} Q_i \right)^{-1/3} \tag{5.6}$$

where α is a fitting parameter, although according to the literature it is defined by

$$\alpha = \left(\frac{48\pi m_e q}{\epsilon_{si} \hbar^2} \right)^{-1/3} \tag{5.7}$$

Clearly the HiSIM model uses the variational method, judging from the form of Equation (5.6). [17] actually makes no mention of the band-gap widening effect, and refers to the change in the charge distribution as being the main quantum effect in the inversion layer. It should be noted that use of Equation (5.6) means that the gate oxide capacitance is no longer constant, but instead is a function of the body and channel charges, and hence of the surface potential. Thus the effective gate oxide width would need to be recalculated for each DC operating point.

The SP (now PSP) model focuses on the other main effect, that of the band gap widening, and ignores the effective oxide width change. The change in the band gap is handled through the following semi-empirical expression [18]

$$\Delta E_g = k_Q \left[\frac{(V_{GB} - V_{FB} - \psi_s)}{t_{ox}^2 T} \right]^{1/3} \tag{5.8}$$

where k_Q is a fitting parameter, and T is the temperature. The term in the brackets effectively gives a normalised expression for the total charge in strong inversion, which is where quantum effects are most pronounced. Having determined ΔE_g, the corresponding change in the surface potential is derived by treating ΔE_g as a small correction to the standard surface potential expression, and then linearising the result.

5.4 New Quantum Mechanical Model

When developing the quantum model for STAG3, it was decided to include both of the main quantum effects associated with confinement of the inversion charge. In other words, two corrections will be made: one to the surface potential to reflect the increase in the band gap, and one to the effective gate oxide thickness to reflect the displacement of the inversion charge away from the silicon-oxide interface. The intention is to provide a degree of flexibility in how quantum effects are modelled, whilst at the same time keeping as much of the basic physics intact as possible.

We have already mentioned that quantum effects only start to have a large effect in or close to strong inversion. Furthermore, one feature of the variational treatment is that it works well in the electrical quantum limit (i.e. high electric field, low temperature), which corresponds more closely to the strong inversion condition. Let us begin with the definition of ΔE_g given in Equations (5.3) and (5.4). In strong inversion, it is valid to write

$$\Delta V \simeq \Delta V_g = \frac{\Delta E_g}{q} \tag{5.9}$$

If we now substitute Equations (5.3) and (5.4) into Equation (5.9), and simplify the q_c factor, we obtain the following

$$\Delta V \simeq \left(\frac{8\hbar^2}{m_e q \epsilon_{si}}\right)^{1/3} \left[q_b + \frac{1}{3}q_c\right]^{2/3} \tag{5.10}$$

We can compare this simple model to the SP model, as in Figure 5.4. We can see that good matching is obtained in the strong inversion region, but that the difference increases below threshold. This behaviour of the SP model is qualitatively correct; the more robust equations in [18] causes ΔV to tend to zero below threshold. In fact, the way that the surface potential model for STAG3 has been constructed means that no similar measures are even necessary. We simply apply our definition of ΔV directly to the low field surface potential ψ_{st}, which we defined in Equation (2.55).

$$\psi_{stqm0} = \psi_{st0} + \Delta V \tag{5.11}$$

$$\psi_{stqmL} = \psi_{stL} + \Delta V \tag{5.12}$$

Now, we recall that we have defined a high field saturation surface potential ψ_{stLsat} in Equation (4.1). What is interesting is that the inclusion of mobility and velocity saturation has resulted in a modified surface potential outside of the strong inversion region only, whereas our quantum treatment is intended to only give an accurate results inside strong inversion. Thus the two sections of our surface potential curve have each been modified by a separate high field effect, and all that remains is to combine them into a single high field expression. We do this by simply taking the smooth minimum of the two curves

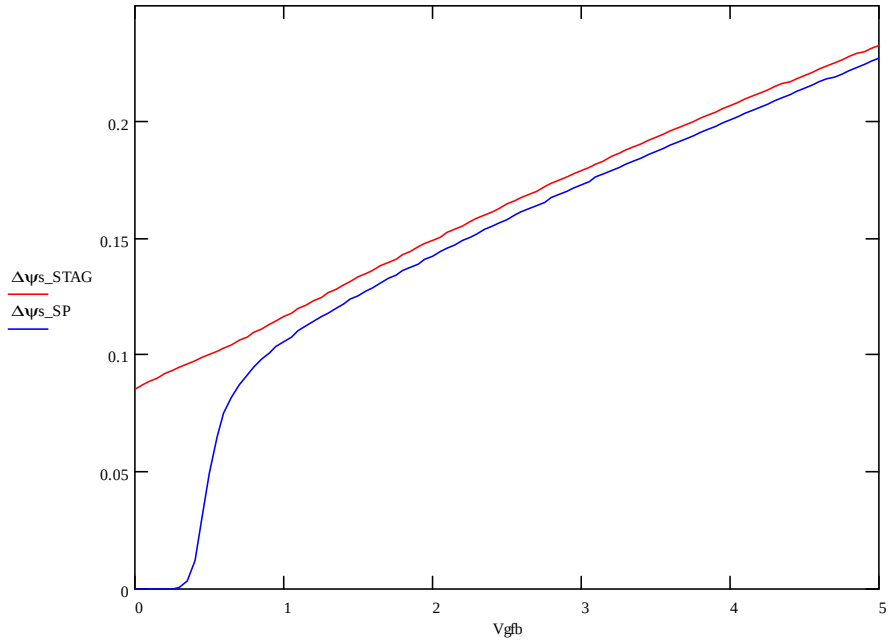

Figure 5.4: *Comparison of ΔV for the STAG3 and PSP models. Although there is close matching between the two expressions when equivalent parameter values are used, PSP displays more physical behaviour in that ΔV moves towards zero outside of the strong inversion regime.*

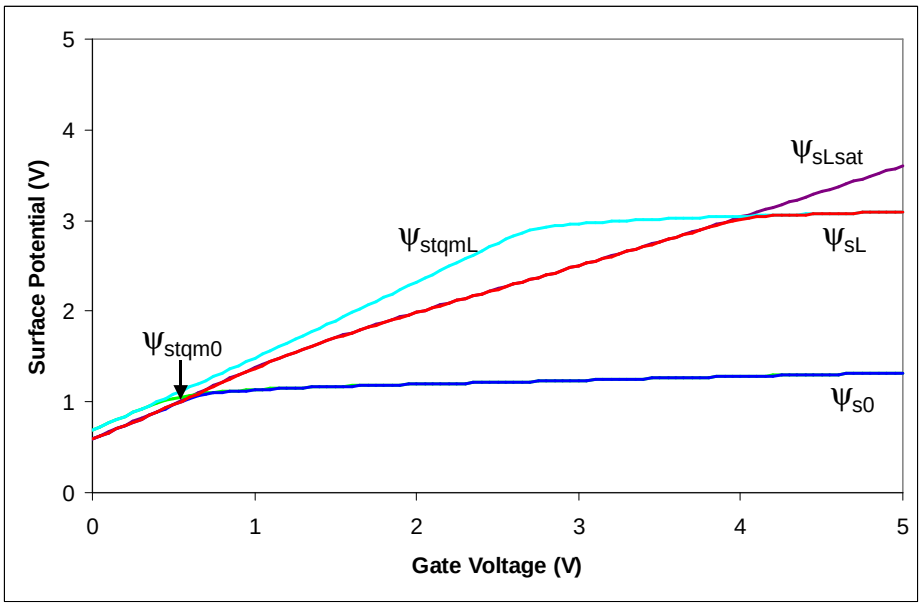

Figure 5.5: *Graphical plot of the mathematical transformations used to obtain ψ_{s0} and ψ_{sL} ($V_{DS} = 1.8V$)*

$$\psi_{s0} = \psi_{sLsat} - 0.5\left[(\psi_{sLsat} - \psi_{stqm0}) + \sqrt{(\psi_{sLsat} - \psi_{stqm0})^2 + 4\phi_t^2}\right] \tag{5.13}$$

$$\psi_{sL} = \psi_{sLsat} - 0.5\left[(\psi_{sLsat} - \psi_{stqmL}) + \sqrt{(\psi_{sLsat} - \psi_{stqmL})^2 + 4\phi_t^2}\right] \tag{5.14}$$

This mathematical transformation is indicated in Figure 5.5. Having derived the high field values for the source and drain surface potential, we take Equation (5.15), and replace μ_0 with μ_{eff}, ψ_{st0} with ψ_{s0}, and ψ_{stL} with ψ_{sL}, to give the final expression for the high field drain current.

$$I_{CH} = \frac{W}{L}\mu_{eff}C_{of}(f(\psi_{sL}) - f(\psi_{s0})) \tag{5.15}$$

References

[1] Frank Stern, and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit", *Physical Review*, vol. 163, no. 3, pp. 816–835, 1967.

[2] Frank Stern, "Self-Consistent Results for n-Type Si Inversion Layers", *Physical Review B*, vol. 5, no. 12, pp. 4891–4899, 1972.

[3] Maarten J. van Dort, Pierre H Woerlee, Andrew J Walker, Casper A. H. Juffermans, and Herbert Lifka, "Influence of High Substrate Doping Levels on the Threshold Voltage and the Mobility of Deep-Submicrometer MOSFETs", *IEEE Transactions on Electron Devices*, vol. 39, no. 4, pp. 932–938, 1992.

[4] M. J. van Dort, P. H. Woerlee, and A. J. Walker, "A Simple Model for Quantisation Effects in Heavily-Doped Silicon MOSFETs at Inversion Conditions", *Solid State Electronics*, vol. 37, no. 3, pp. 411–414, 1994.

[5] G. Chindalore, S. A. Hareland, S. Jallepalli, A. F. Tasch, Jr., C. M. Maziar, V. K. F. Chia, and S. Smith, "Experimental Determination of Threshold Voltage Shifts Due to Quantum Mechanical Effects in MOS Electron and Hole Inversion Layers", *IEEE Electron Device Letters*, vol. 18, no. 5, pp. 206–208, 1997.

[6] J. R. Brews, "A charge-sheet model of the MOSFET", *Solid State Electronics*, vol. 21, pp. 345–355, 1978.

[7] Yutao Ma, Litian Liu, Zhiping Yu, and Zhijian Li, "Simplified Method to Investigate Quantum Mechanical Effects in MOS Structure Inversion Layer", *IEEE Transactions on Electron Devices*, vol. 47, no. 6, pp. 1303–1305, 2000.

[8] T. Ando, A. B. Fowler, and F. Stern, "Properties of 2-dimensional electron system", *Reviews in Modern Physics*, vol. 54, pp. 437–672, 1982.

[9] Y. Ma, Z. Li, L. Liu, L. Tian, and Z. Yu, "Effective density-of-states approach to QM correction in MOS structures", *Solid State Electronics*, vol. 44, pp. 401–407, 2000.

[10] M. Rao, and A. Rothworf "Variational calculations of ground and higher states of metal-oxide-semiconductor inversion layer in a nonuniform electric field", *Journal of Applied Physics*, vol. 73, no. 7, pp. 3385–3389, 1992.

[11] Mhammad Z. Kauser, Md S. Hasan, and Anisul Haque, "Effects of Wave Function Penetration Into the Gate-Oxide on Self-Consistent Modeling of Scaled MOSFETs", *IEEE Transactions on Electron Devices*, vol. 49, no. 4, pp. 693–695, 2002.

[12] Yutao Ma, Litian Liu, Zhiping Yu, and Zhijian Li, "Validity and Applicability of Triangular Potential Well Approximation in Modelling of MOS Structure Inversion and Accumulation Layer", *IEEE Transactions on Electron Devices*, vol. 47, no. 9, pp. 1764–1767, 2000.

[13] Fabien Pregaldiny, Christophe Lallement, Ronald van Langevelde, and Daniel Mathiot, "An advanced explicit surface potential model physically accounting for the quantization effects in deep-submicron MOSFETs", *Solid State Electronics*, vol. 48, pp. 427–435, 2004.

[14] Fabien Pregaldiny, Christophe Lallement, and Daniel Mathiot, "Accounting for quantum mechanical effects from accumulation to inversion, in a fully analytical surface-potential-based MOSFETs model", *Solid State Electronics*, vol. 48, pp. 781–787, 2004.

[15] M.I. Vexler, "A simple quantum model for the MOS strcuture in accumulation mode", *Solid State Electronics*, vol. 47, pp. 1283–1287, 2003.

[16] R. Clerc, P. O'Sullivan, K. G. McCarthy, G. Ghibaudo, G. Pananakakis, and A. Mathewson, "A physical compact model for direct tunneling from NMOS inversion layers", *Solid State Electonics*, vol. 45, pp. 1705–1716, 2001.

[17] M. Miura-Mattausch, H. Ueno, H. J. Mattausch, S. Kumashiro, T. Yamaguchi, K. Yamashita, and N. Nakayama, "Circuit Simulation Models for Coming MOSFET Generations", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E85-A, no. 4, pp. 740–747, Apr. 2002.

[18] G. Gildenblat, T. L. Chen, and P. Bendix, "Closed-form approximation for the perturbation of MOSFET surface potential by quantum mechanical effects", *Electronics Letters*, vol. 36, no. 12, pp. 1072–1073, 2000.

Chapter 6

Threshold Voltage Model

6.1 Introduction

In a surface potential based model, it is a misnomer to describe part of the model as a 'threshold voltage model', since threshold voltage is a somewhat vague and ill-defined term, and so is not used directly in the model. For this reason, most of the effects described in this chapter are actually handled by modifying the silicon body factor γ_s, which in turn modifies the threshold voltage. That being said, all of the model improvements detailed in this chapter manifest themselves primarily through a shift in the threshold voltage, and this term is so widely used that it makes sense to lump them under that heading. The model improvements dealt with in this chapter are: the effect of non-uniform doping in the vertical direction, the short channel effect, the reverse short channel effect, and drain induced barrier lowering.

There are of course a number of other effects which will modify the threshold voltage and yet are not included in this chapter. They include the polysilicon depletion effect, and quantum effects. Early theoretical models generally used a threshold voltage shift to express the impact of these effects on the device behaviour [1,2]. Modern compact models normally handle these effects by modifying the surface potential [3,4]. As we have seen in Chapters 2 and 5, this more modern approach has been adopted in STAG3.

The chapter concludes with a detailed look at the way in which the threshold voltage parameter ties in with the core surface potential model. While the threshold voltage can defined in a variety of ways, and extracted using numerous different methods, we show that it is possible to quantify and minimise the degree of error involved when relating this quantity to the well-defined flat band voltage. In this way, we are better able to bridge the gap between the empirical models frequently used in characterisation and circuit design, and the more strictly defined models used in modern compact models.

6.2 Modelling of Non-Uniform Doping

Modern MOSFET devices are typically doped by means of ion implantation. It is usually valid to assume that the result of this and the subsequent high-temperature annealing stages is a silicon layer containing dopants in an approximately Gaussian distribution [5]. We have chosen to model this distribution as a combination of two components; a uniform background doping concentration, and a varying part which follows the Gaussian distribution

$$N(x) = N_0 + (N_{\max} - N_0) \exp\left(\frac{(x - R_p)^2}{2\Delta R_p^2}\right) \tag{6.1}$$

where N_0 is the uniform background dopant concentration, N_{max} is the maximum doping concentration, which occurs at $x = R_p$, x is the depth into the silicon, measured from the oxide-silicon interface, R_p is the projected range of the doping implant (modified by any subsequent processing stages), and ΔR_p is the standard deviation. Thus, any such profile can be characterised by 4 parameters: N_0, N_{max}, R_p, and ΔR_p. In Figure 6.1 we show two different profiles corresponding to different values of these 4 parameters.

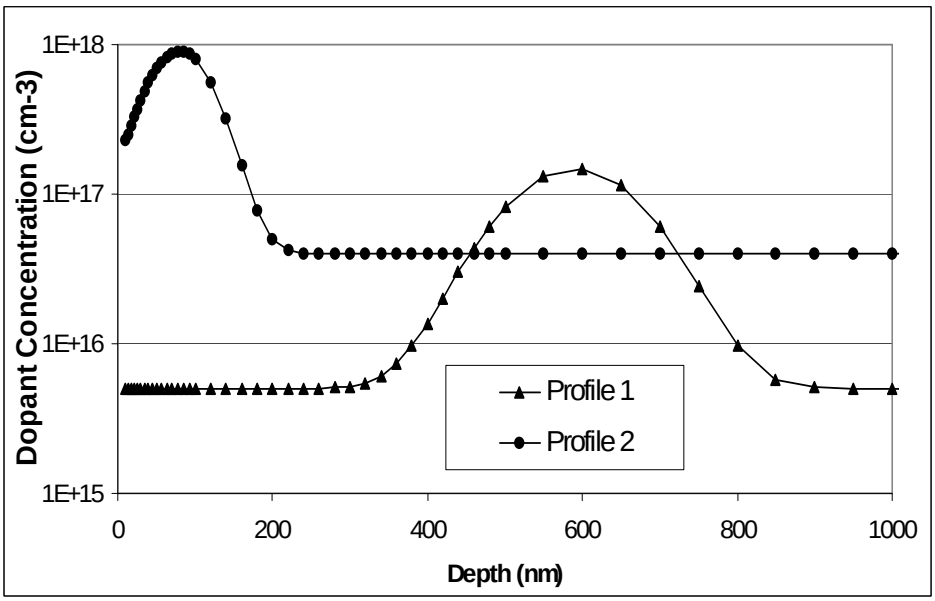

Figure 6.1: *Two example dopant profiles. Profile 1:* $R_p = 100nm$, $\Delta R_p = 590nm$, $N_{max} = 2 \times 10^{17} cm^{-3}$, $N_0 = 5 \times 10^{15} cm^{-3}$. *Profile 2:* $R_p = 80nm$, $\Delta R_p = 40nm$, $N_{max} = 9 \times 10^{17} cm^{-3}$, $N_0 = 4 \times 10^{16} cm^{-3}$.

The first curve corresponds to the case where $R_p \gg \Delta R_p$. In practise this condition is met when $R_p > 3\Delta R_p$. In this case, the doping concentration at the silicon-oxide interface is equal to N_0. We can identify three distinct regions within the silicon layer, as indicated on Fig. 6.2. We can imagine that as the surface potential ψ_s (effectively the gate-body voltage) is increased, the edge of the depletion region, which we shall denote by x_d, passes

through each of these three regions in turn. Region 1 is identical to a silicon layer with uniform doping N_0, and the relation between x_d and ψ_s is given by the standard expression relating depletion depth to applied voltage

$$\psi_{s1} = \frac{qN_0x_d^2}{2\epsilon_{si}} \tag{6.2}$$

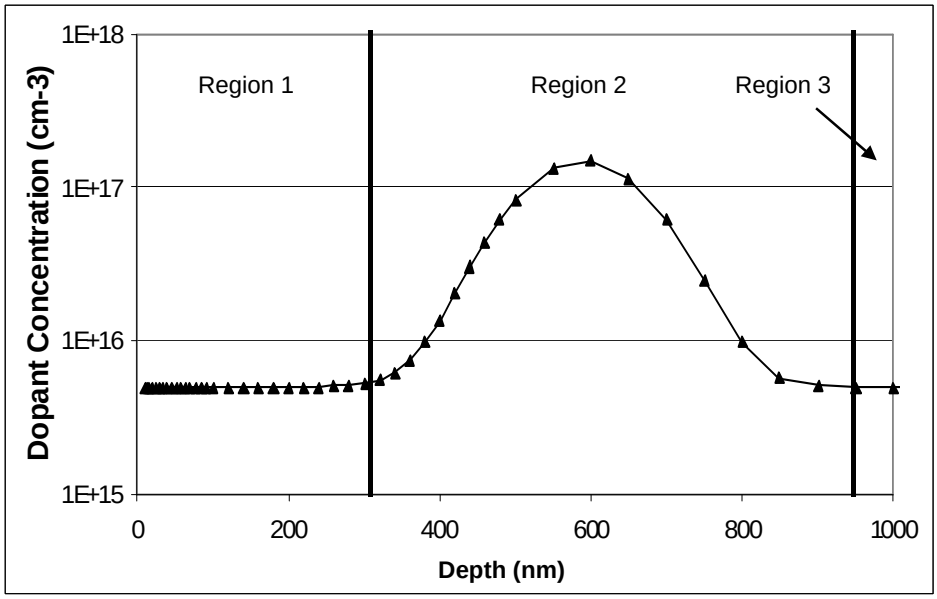

Figure 6.2: *Profile 1, showing 3 distinct modelling regions.*

Once x_d reaches the edge of the non-uniform profile (Region 2), things get more complicated. The mathematical derivation for this region is somewhat involved; here we simply present the result

$$\psi_{s2} = \frac{q}{\epsilon_{si}}\left(\frac{N_0x_d^2}{2} + (N_{\max} - N_0)\,\Delta R_p\left[\sqrt{\frac{\pi}{2}}R_pR_{erf} + R_{exp}\right]\right) \tag{6.3}$$

where

$$R_{erf} = \Delta R_p\left[erf\left(\frac{x_d - R_p}{\sqrt{2}\Delta R_p}\right) + erf\left(\frac{R_p}{\sqrt{2}\Delta R_p}\right)\right] \tag{6.4}$$

and

$$R_{exp} = \Delta R_p\left[\exp\left(-\frac{R_p^2}{2\Delta R_p^2}\right) - \exp\left(\frac{(x_d - R_p)^2}{2\Delta R_p^2}\right)\right] \tag{6.5}$$

Note that erf is the error function. We will return to this result several times over the course of the chapter. For now, suffice to say that it is not possible to rearrange (6.3) such that x_d can be obtained in closed-form as a function of ψ_s. Of course, this is trivial for

(6.2) (uniform doping case).

Continuing on to Region 3, we see that we can obtain a simplified expression by applying the condition $x_d \gg R_p$ to (6.3). We then obtain

$$\psi_{s3} = \frac{q}{\epsilon_{si}} \left[\frac{N_0 x_d^2}{2} + (N_{\max} - N_0)\, \Delta R_p \left(\sqrt{\frac{\pi}{2}} R_p \left[erf\left(\frac{R_p}{\sqrt{2}\Delta R_p} \right) + 1 \right] + \Delta R_p \exp\left(-\frac{R_p^2}{2\Delta R_p^2} \right) \right) \right] \quad (6.6)$$

Before we continue it should be noted that the relation (6.3) governs the behaviour of the depletion region in all three regions. However, in Regions 1 and 3, it is possible to make asymptotic simplifications such that x_d can be expressed as a closed-form function of ψ_s. It is only in Region 2 that no such simplification is possible.

The second profile, shown in Fig. 6.3, corresponds to a case where R_p is comparable to ΔR_p. We can see that Region 1 does not appear, and Regions 2 and 3 are as before. From a modelling point of view, the two types of profile shown in Figs. 6.2 and 6.3 are the only two that matter. For $\Delta R_p \gg R_p$, we find that $N(x) \to N_{max}$, and hence this case is trivial.

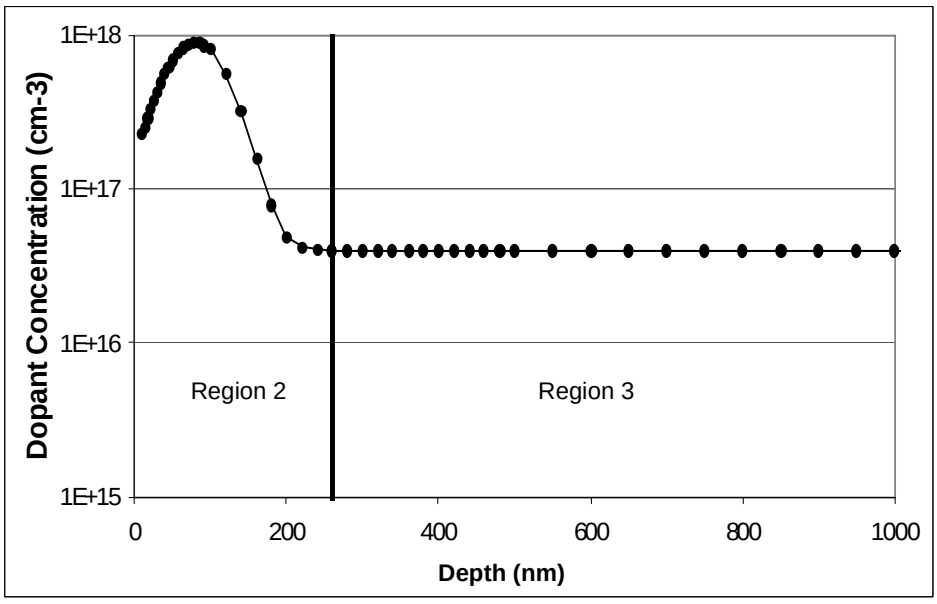

Figure 6.3: *Profile 2, showing 2 distinct modelling regions (no Region 1).*

Our goal is to obtain an accurate closed-form expression which allows us to determine x_d as a closed-form function of ψ_s. However, an examination of (6.3) reveals two major problems. The first is the transcendental nature of the equation. The second is the presence of the error function, itself an iterative function. This complexity makes (6.3) unsuitable for circuit simulation without modification [6]. A number of approximate models are outlined

in [6], many of which rely on breaking the profile into a two region step profile, each with its own constant value of doping concentration. The most accurate of these early models is the doping transformation model [7], but while it gives reasonable results, it fails to capture the true functional dependence of x_d on ψ_s. Furthermore, being a two section model, mathematical discontinuities are introduced when transitioning between the two regions, as the effective body factor jumps from one value to another.

It should be noted that many compact models have attempted to circumvent the issue of modelling Gaussian-type dopant profiles. The EKV model uses an empirical expression for the dopant profile, using a formulation similar to [7], but which removes the discontinuity in the first derivative [8]. However, it is implicitly assumed that the doping concentration is the body is lower than that at the surface, so this model could not be used to simulate a retrograde profile, for instance. The SP compact model, prior to its evolution into the newer PSP model, employed a single empirical expression with a number of fitting parameters to account for the effects of a non-uniform doping profile, as well as normal and reverse short channel effects [9]. This expression was then used to modify the body factor. In contrast, the PSP model uses a somewhat more physical approach; this time an empirical expression is used to approximate the profile itself [10]. However, the expression presented in [10] is specifically intended to simulate a retrograde-type dopant profile, and so it can be considered as having the opposite problem to the EKV model. Although it is stated in [9] that other, more complex expressions can be employed, it is not clear what these are, how much flexibility they provide, and how many additional parameters are required.

6.3 New Non-Uniform Doping Model

6.3.1 Approximation of the Error Function

We shall now propose a new model which accurately models Gaussian-type profiles, without the need for iterative algorithms. Of the two problems outlined in the last section, the presence of the error function is the easiest to solve. Although it is an iterative function, many studies have been made to devise accurate closed-form approximations [11–17]. Many of these studies date back to the 1970s and 1980s, when limited computing power made numerical iterations even less desirable than today. Probably the most suitable approximation is given by Menzel [11, 12]

$$erf(a) \approx \sqrt{1 - \exp\left(-\frac{4a^2}{\pi}\right)} \tag{6.7}$$

This equation gives a maximum percentage error of 0.7 percent for all positive values of a. Clearly we desire similar accuracy for negative values of a ($0 < x_d < R_p$) as well, so we apply a simple modifier and our x_d dependent error function term can now be approximated by

$$erf\left(\frac{x_d - R_p}{\sqrt{2}\Delta R_p}\right) \approx \frac{x_d - R_p}{\sqrt{(x_d - R_p)^2}}\sqrt{1 - \exp\left(-\frac{4}{\pi}\frac{(x_d - R_p)^2}{2\Delta R_p^2}\right)} \tag{6.8}$$

A look at Fig. 6.4 should be enough to convince the reader that this aspect of the problem has been handled to a very high degree of accuracy. In Fig. 6.5 the percentage error has been plotted as a function of the depletion depth, and it can indeed be seen that the error does not exceed 0.7 percent.

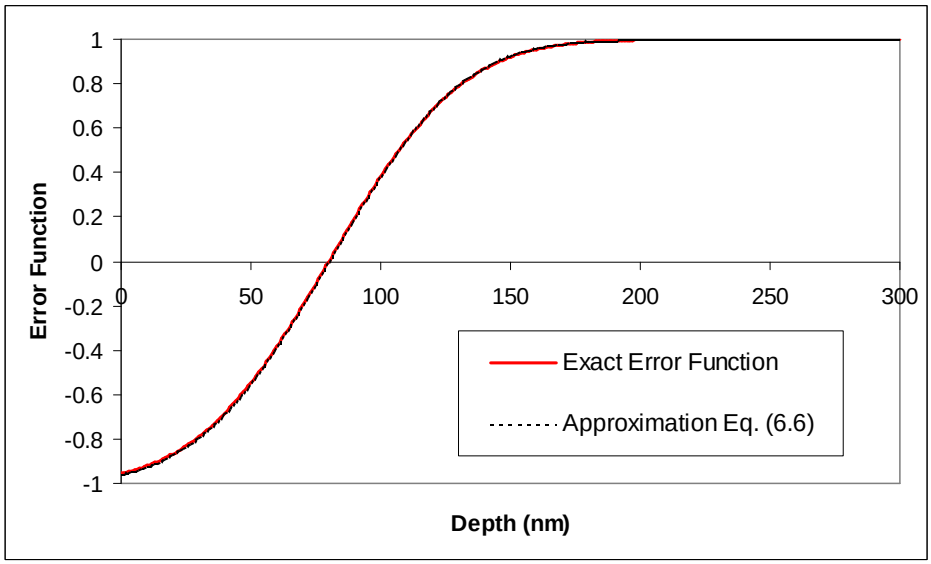

Figure 6.4: *Exact and approximate error functions expressions, plotted vs increasing depletion depth.* $R_p = 80nm$, $\Delta R_p = 40nm$, $N_{max} = 9 \times 10^{17} cm^{-3}$, $N_0 = 4 \times 10^{16} cm^{-3}$.

6.3.2 Obtaining a Closed-Form Expression for x_d

The problem of directly approximating (6.3) with a closed-form expression poses greater difficulties, which is why a different approach will be taken. Let us replace (6.3) with the following equation

$$\psi_s = \frac{qN_{eff}(\psi_s)x_d^2}{2\epsilon_{si}} \tag{6.9}$$

In writing this equation, we are defining an *effective doping concentration* $N_{eff}(\psi_s)$, which can be obtained through some as yet unspecified mathematical transformation of the real doping profile $N(x)$, and which allows us to define x_d as a closed-form function of ψ_s through simple rearrangement

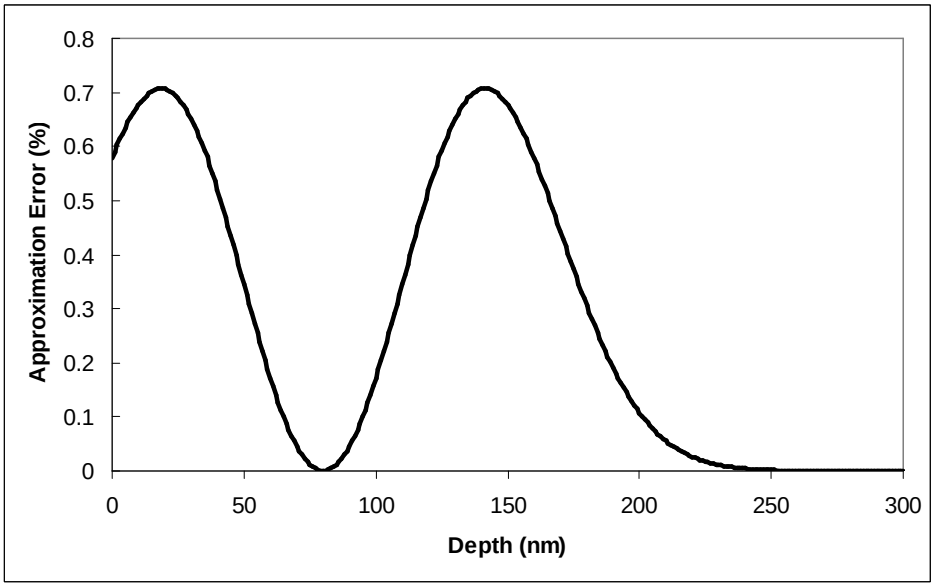

Figure 6.5: *Percentage error of the approximate solution shown in Fig. 6.4.*

$$x_d = \sqrt{\frac{2\epsilon_{si}\psi_s}{qN_{eff}\left(\psi_s\right)}} \tag{6.10}$$

Just as Equation (6.3) provides us with a universal (albeit intractable) relation, Equation (6.10) can be used to model the variation of x_d with ψ_s across all three regions. The key is to somehow determine the correct functional form of $N_{eff}(\psi_s)$. Fortunately, we already have some simplifications to use. Comparison of Equations (6.2) and (6.9) shows that $N_{eff}(\psi_s) = N_{eff1} = N_0$ in Region 1. In Region 3, we can rearrange Equation (6.6) to give

$$x_{d3} = \sqrt{\frac{2}{N_0}\left[\frac{\epsilon_{si}\psi_s}{q} - \left(N_{\max} - N_0\right)\Delta R_p\left(\sqrt{\frac{\pi}{2}}R_p\left[erf\left(\frac{R_p}{\sqrt{2}\Delta R_p}\right)+1\right] + \Delta R_p \exp\left(-\frac{R_p^2}{2\Delta R_p^2}\right)\right)\right]} \tag{6.11}$$

Substituting x_{d3} for x_d in (6.9) and rearranging gives

$$N_{eff}\left(\psi_s\right) \rightarrow N_{eff3}\left(\psi_s\right) = \frac{2\epsilon_{si}\psi_s}{qx_{d3}^2} \tag{6.12}$$

This leaves us the with the task of finding a suitable form for N_{eff} in the problematic Region 2. Rather than attempt any formal mathematical transformation, which would leave us in no better a position than before, we will opt for an empirical fitting approach instead. We start by defining our profile, by determining appropriate values of N_0, N_{max}, R_p, and ΔR_p. The next step is to define a suitable range of values for x_d, and then using a

mathematical software simulation tool (Mathcad 7.0 was used for this work) to calculate the corresponding values of ψ_s, using Equation (6.3). Having done this we can express N_{eff} as a function of x_d easily enough by rearranging (6.9).

$$N_{eff}(\psi_s) = \frac{2\epsilon_{si}\psi_s}{qx_d^2} \tag{6.13}$$

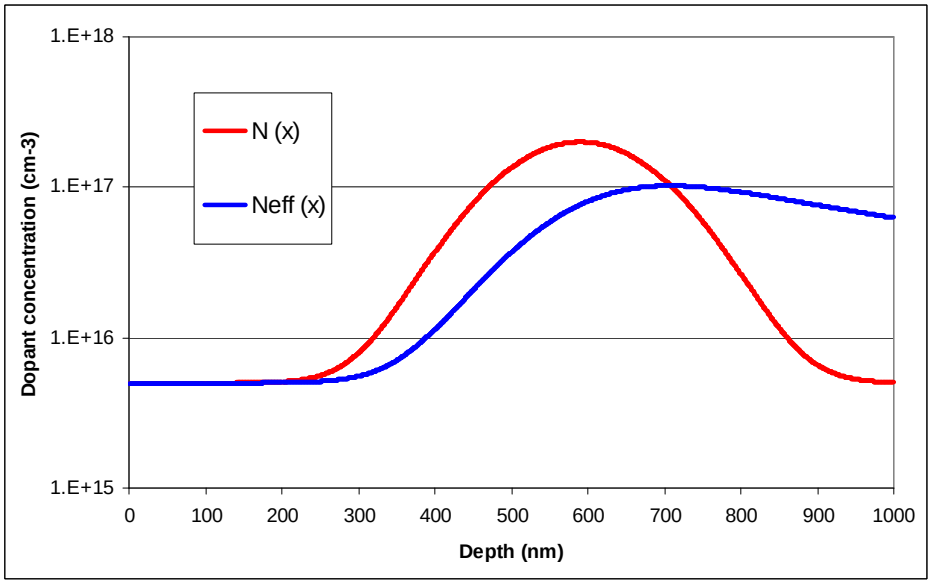

Figure 6.6: *True doping profile, and effective doping profile corresponding to Equation (6.9), for Profile 1.*

Between Equation (6.13) and the implicit relation given in Equation (6.3), we now have enough information to plot N_{eff} as a function of ψ_s, using our mathematical software. Figs. 6.6 and 6.7 show our two profiles, with both the true doping profile N and the effective profile N_{eff} plotted as a function of x_d. In both cases, we have indicated the approximate boundaries marking Region 2, the area in which we need to achieve an accurate approximation to N_{eff}. Notice that we have not selected the whole of Region 2 in the case of $\Delta R_p = 50$nm. It has been found that fitting the curve is more difficult when ΔR_p has a mid-range value, since Region 2 accounts for a larger range of ψ_s values than when $R_p \gg \Delta R_p$. We therefore avoid trying to fit the part of Region 2 close to the silicon-oxide interface, since this will be accounted for when we construct our complete closed-form expression for x_d. Also, it should be appreciated that very high levels of accuracy are not required. This is because, as can be seen in Equation (6.10), x_d is dependent on the square root of N_{eff}, which serves to mitigate any errors in our approximation.

Looking at the selected parts of each curve, it seems that a polynomial approximation might suffice. Remembering that the aim is to fully define N_{eff} as a closed form function of ψ_s, we can test this idea with the following equation

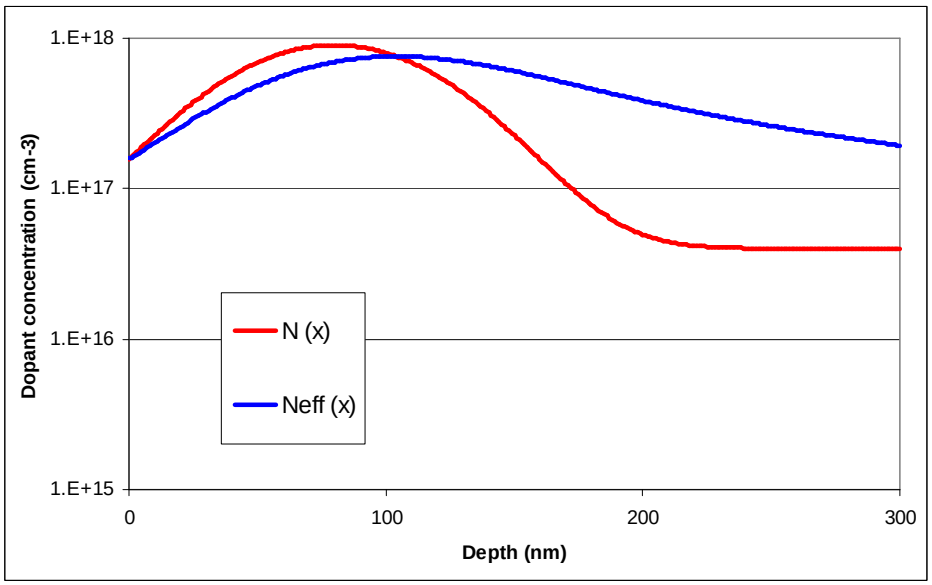

Figure 6.7: *True doping profile, and effective doping profile corresponding to Equation (6.9), for Profile 2.*

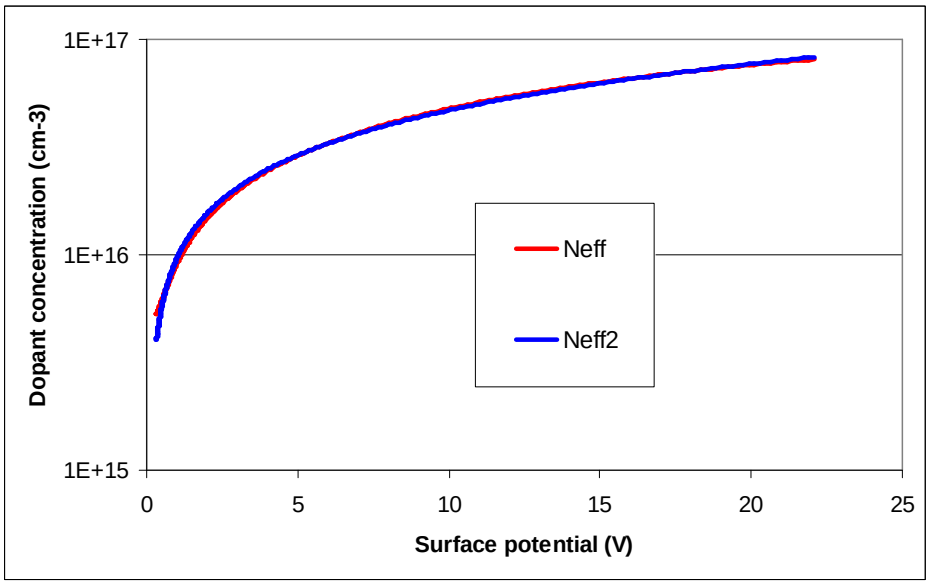

Figure 6.8: *Matching of polynomial expression to effective doping profile (Profile 1).*

$$N_{eff2} = a_{x_d} + b_{x_d}\psi_s^{1/2} + c_{x_d}\psi_s \tag{6.14}$$

where a_{x_d}, b_{x_d}, and c_{x_d} are fitting parameters. The x_d subscript reminds us that these parameters relate to our closed-form expression for the depletion depth x_d as a function of ψ_s. Again, using our mathematical software environment, it is a straightforward procedure

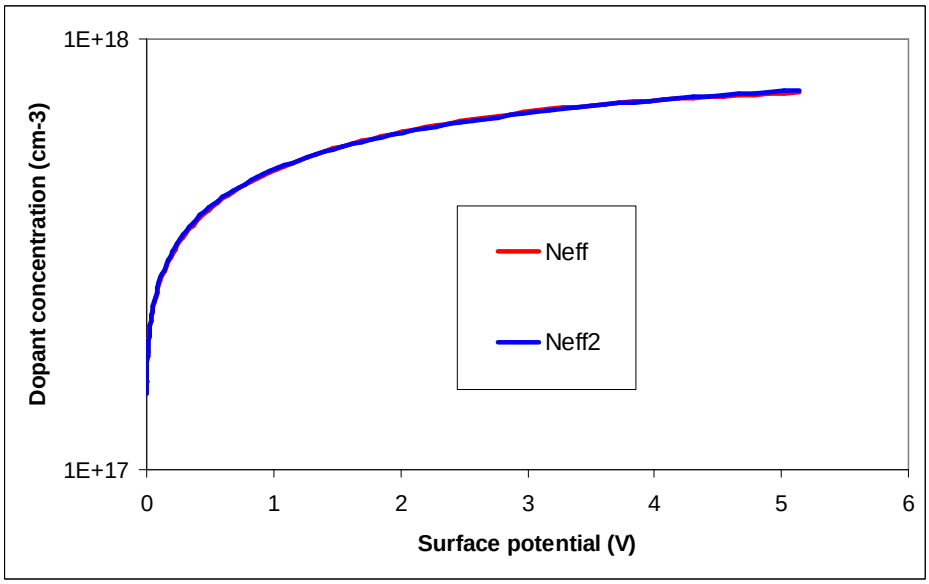

Figure 6.9: *Matching of polynomial expression to effective doping profile (Profile 2).*

to obtain best fit values for these three parameters for any given range. The best fit cases for our two profiles are also indicated on Figs. 6.8 and 6.9. It can be seen that good fits are obtained for both curves, with a particularly close fit in the case of $\Delta R_p = 10$nm. Having demonstrated that Equation (6.14) can be used to approximate N_{eff2}, we can now combine it with the expressions for N_{eff1} and N_{eff3} to obtain a complete expression for N_{eff} across the entire range of ψ_s. This is achieved through the following expressions

$$N_{eff23} = N_{eff2} - 0.5\left[(N_{eff2} - N_{eff3}) + \sqrt{(N_{eff2} - N_{eff3})^2 + 4\epsilon_{23}^2}\right] \tag{6.15}$$

$$N_{effapprox} = N_{eff23} - 0.5\left[(N_{eff23} - N_0) - \sqrt{(N_{eff23} - N_0)^2 + 4\epsilon_{12}^2}\right] \tag{6.16}$$

Figs. 6.10 and 6.11 show the matching between $N_{effapprox}$ and N_{eff} for the two profiles. N_{eff2} and N_{eff3} are also shown for completeness. Overall the fit is good, with the most pronounced deviations occuring well away from the low surface potential region corresponding to actual device operation ($V_{DS} = 0.1V$). Thus we can see that a suitable choice of polynomial fitting constants is sufficient to give good fitting.
Finally, having obtained a suitable approximation for N_{eff}, we substitute $N_{effapprox}$ for N_{eff} in Equation (6.10), to obtain the final expression for x_d

$$x_d = \sqrt{\frac{2\epsilon_{si}\psi_{st0}}{qN_{eff0}}} \tag{6.17}$$

where

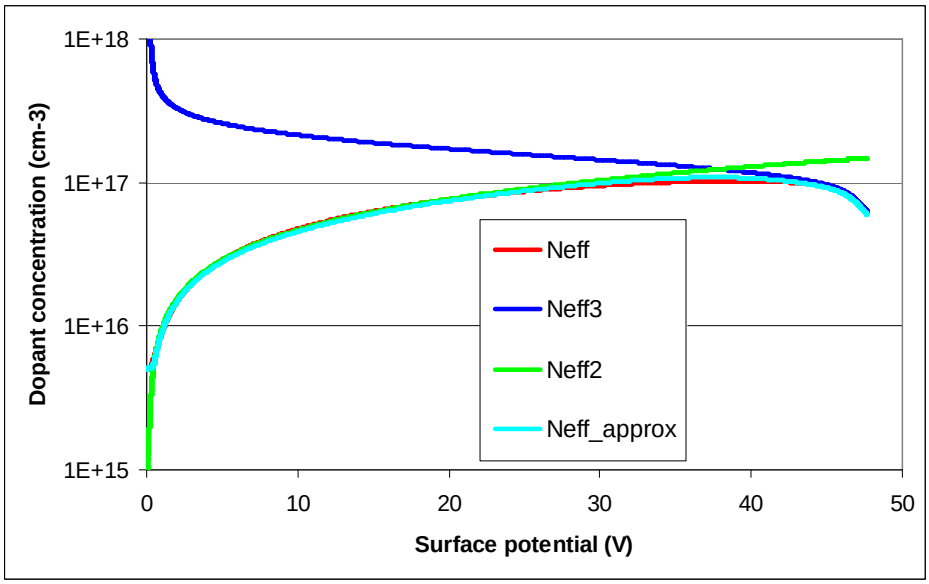

Figure 6.10: *Comparison of exact and approximate effective doping profiles for Profile 1.*

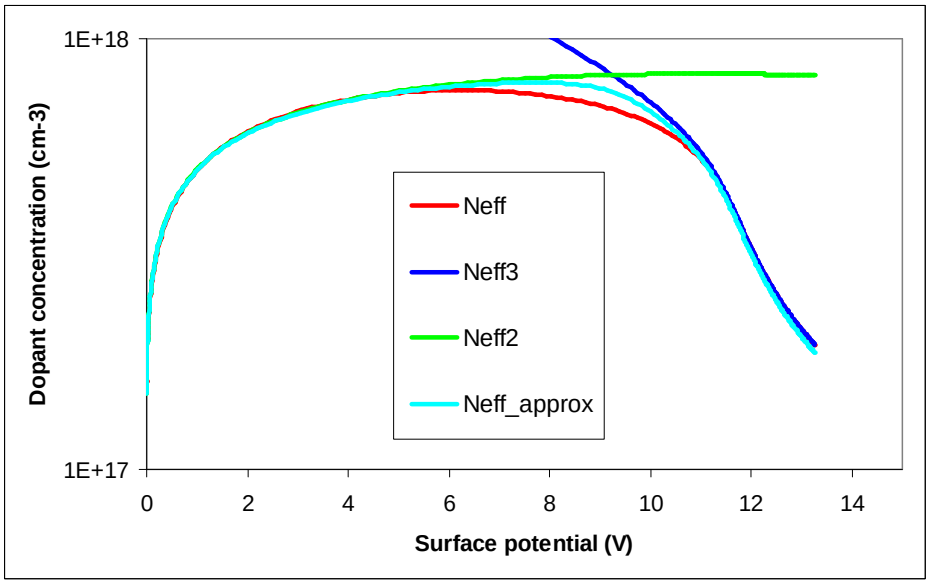

Figure 6.11: *Comparison of exact and approximate effective doping profiles for Profile 2.*

$$N_{eff0} = N_{effapprox}(\psi_{st0}) \tag{6.18}$$

Figs. 6.12 and 6.13 plots the approximate relation between x_d and ψ_s, and compares it to the exact relation given by Equations (6.3)-(6.5). This time we are restricting the plot to low values of ψ_s, which is the relevant region for real device operation. Fitting is very good, and confirms that the approximation method does indeed work to provide a set of accurate, analytic expressions which reproduce the effects of a non-uniform doping profile.

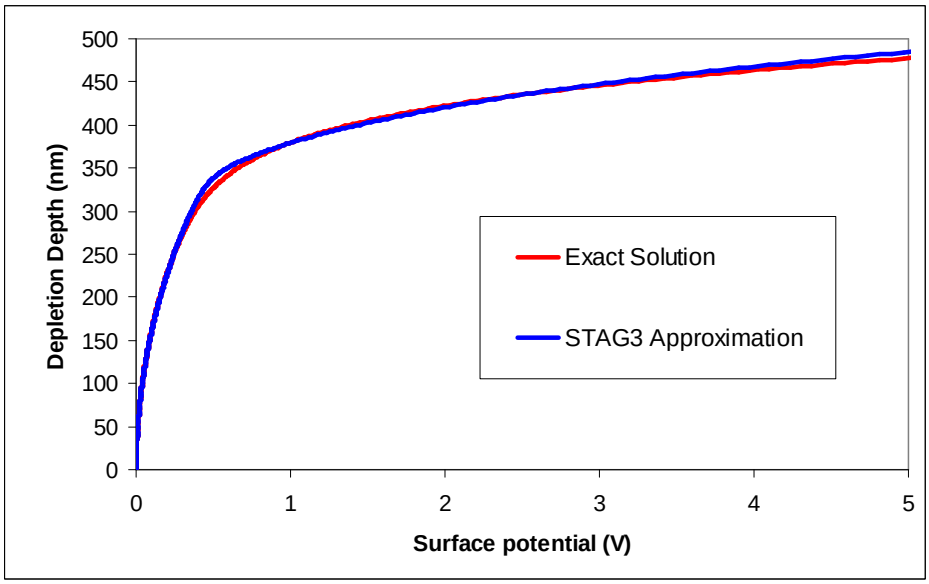

Figure 6.12: *Comparison of exact and approximate x_d-ψ_s relation for Profile 1.*

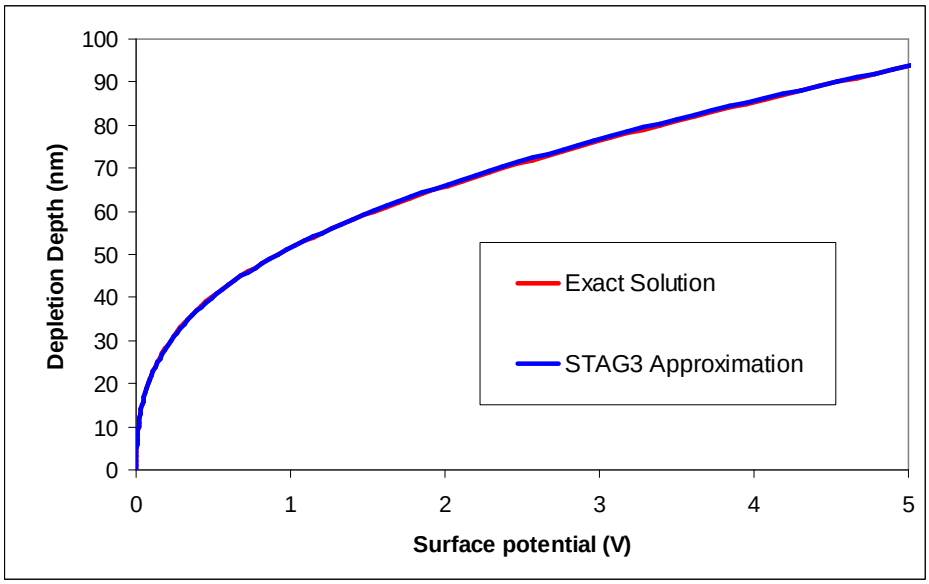

Figure 6.13: *Comparison of exact and approximate x_d-ψ_s relation for Profile 2.*

6.3.3 Modifying the Body Factor γ_s

Since we can now express x_d as a closed-form function of the surface potential, and hence of the terminal voltages, we are now in a position to calculate the total depleted body charge Q_b for a non-uniform doping profile under any set of bias conditions.

$$Q_b = -q \int_0^{x_d} N(x)dx \tag{6.19}$$

And the expression for threshold voltage in the case of a non-uniformly doped device is [6]

$$V_{TH} = V_{FB} + \psi_{si} - \frac{q}{C_{ox}} \int_0^{x_d} N(x)dx \tag{6.20}$$

From Equation (6.1) we obtain

$$\int_0^{x_d} N(x)dx = N_0 x_d + (N_{\max} - N_0)\sqrt{\frac{\pi}{2}}\Delta R_p \left[erf\left(\frac{x_d - R_p}{\sqrt{2}\Delta R_p}\right) - erf\left(\frac{-R_p}{\sqrt{2}\Delta R_p}\right)\right] \tag{6.21}$$

$$= N_0 x_d + (N_{\max} - N_0)\sqrt{\frac{\pi}{2}}\Delta R_p \left[erf\left(\frac{x_d - R_p}{\sqrt{2}\Delta R_p}\right) + erf\left(\frac{R_p}{\sqrt{2}\Delta R_p}\right)\right] \tag{6.22}$$

Substituting Equations (6.17) and (6.22) into Equation (6.20) gives

$$V_{TH} = V_{FB} + \psi_{si} + \frac{qN_0}{C_{of}}\sqrt{\frac{2\epsilon_{si}\psi_{si}}{qN_{eff}\left(\psi_{si}\right)}} + \frac{q\left(N_{\max} - N_0\right)}{C_{of}}\sqrt{\frac{\pi}{2}}R_{erf} \tag{6.23}$$

While this is an acceptable way of modifying the classic threshold voltage relation, we would prefer to express the effect of non-uniform doping by directly modifying the silicon body factor γ_s. If we look at the definition of γ_s in Equation (2.14), we can see that it is possible to re-write Equation (6.23) in the following way:

$$V_{TH} = V_{FB} + \psi_{si} + \gamma_s\left(\sqrt{\frac{N_0}{N_{eff}\left(\psi_s\right)}}\psi_{si} + \left(N_{\max} - N_0\right)\sqrt{\frac{\pi q}{4\epsilon_{si}N_0}}R_{erf}\right) \tag{6.24}$$

Before we leave this section, a brief discussion concerning parameter extraction is in order. While great care has been taken to ensure that the STAG3 non-uniform doping model can be used with any generic Gaussian-type profile, we have seen that there will always be some part of the profile which is modelled with less accuracy than the others. When extracting the three fitting parameters, it is important to consider the type of profile, and to have some idea of what range of values x_d can be expected to have, based on the operating voltages of the device. In this way, greater accuracy can be achieved in the range of interest.

For instance, it is common for MOSFETs to receive a deep high dose implant, in order to ensure that punch-through and short-channel effects are suppressed, whilst at the same time keeping the surface doping concentration relatively low, so that the carrier mobility is not adversely affected by Coulomb scattering from ionised impurities. This would correspond to a profile with a high value of N_{max} relative to N_0, a large value of R_p, and perhaps a relatively low value of ΔR_p. It might be found that under normal operating conditions, the edge of the depletion region passes through the uniformly-doped Region 1, only reaching Region 2 at around the onset of strong inversion. Beyond that point, the relatively slow increase of x_d with respect to ψ_s (due to screening from the inversion charge), combined with the increased doping, might mean that the depletion region does not actually penetrate very far into Region 2. In that case, the best accuracy can be achieved by fitting

parameters to that part of Region 2 which can be depleted, and ensuring an accurate transition between Regions 1 and 2. The rest of the profile can be largely ignored, though of course we are always guaranteed good accuracy once the depletion region reaches Region 3.

If the same profile were to be moved much closer to the silicon-oxide interface, so that now R_p is small, we might find that the entire profile has to be accounted for, and so our approximation would become more difficult. If, for instance, we place less emphasis on approximating the profile close to the interface, the likely result is that we will get poorer accuracy in the subthreshold region. But in any case, it has been shown that the relative error associated with the Region 2 approximation can be kept quite small, and certainly represents a significant improvement on other approximate models such as those employing step profiles.

6.4 Drain Induced Barrier Lowering

The drain induced barrier lowering (DIBL) effect is frequently modelled via an empirical threshold shift relation [6]. As the STAG model is physically based, an effective flat-band voltage is introduced instead

$$V^f_{FB\text{eff}} = V^f_{FB} - \frac{\sigma}{L} \cdot V_{DS} \tag{6.25}$$

6.5 Short Channel Effect

The well known short-channel and narrow width effects are modelled via a modified body factor [6]

$$\gamma_{\text{eff}} = \gamma_s \cdot \left(1 - \frac{\Delta L}{L}\right) \cdot \left(1 + \frac{\Delta W}{W}\right) \tag{6.26}$$

where ΔL and ΔW are left as model parameters. The modified γ_{eff} is then used throughout in the evaluation of the surface potentials. This is identical to the STAG2 equation.

6.6 Reverse Short Channel Effect

In addition to the standard short channel effect described above, it has been found that for some devices, the threshold voltage displays the opposite behaviour. As the channel length is decreased, the threshold voltage first increases, before decreasing again in the usual manner. The result is a hump in the threshold voltage versus channel length characteristics. This behaviour is known as the anomalous short channel effect, or the reverse short channel effect (RSCE) [18, 19].

There is strong evidence that the reverse short channel effect is caused by a build-up of active channel dopant close to the source and drain [20–22]. This is a common occurrence in sub-micron devices, and has its origin in the high energy implants used to define the

source and drain junction regions. The implants cause substantial damage where they are made, and this region of damage acts as a source of silicon intersititials during subsequent high temperature annealing stages. It has been shown that these interstitials act as catalysts which promote diffusion of dopant ions towards the surface. As a result, the dopant concentration is greatest around the source and drain regions, and decreases towards the centre of the channel.

The effect of this dopant pile-up at each end of the channel causes an effective increase in the average channel doping concentration. The smaller the length of the channel, the more pronounced the effect will be, hence we see an increase in the threshold voltage as L decreases. In order to model this effect, we need to calculate N_{avg}, the averaged doping concentration, and then use this in Equation (6.20).

$$N_{avg} = \frac{1}{L}\int_0^L N(y)\,dy \tag{6.27}$$

A common assumption when modelling the RSCE is that the dopant profile decreases exponentially from the source and drain ends of the channel [23, 24]. We therefore can express Equation (6.27) in the following form

$$N_{avg} = \frac{1}{L}\int_0^L N_{ch} + (N_{rsce}(y) - N_{ch})\left[\exp\left(-\frac{y}{L_{rsce}}\right) + \exp\left(\frac{y-L}{L_{rsce}}\right)\right] dy \tag{6.28}$$

where N_{ch} is the position independent background component of the channel doping, N_{rsce} is the maximum value of the position dependent component associated with the RSCE, and L_{rsce} is the characteristic length associated with the doping profile. Solving Equation (6.28), we obtain

$$N_{avg} = N_{ch}\left[1 + 2\frac{L_{rsce}}{L}\left(\frac{N_{rsce}(y)}{N_{ch}} - 1\right)\left(1 - \exp\left[-\frac{L}{L_{rsce}}\right]\right)\right] \tag{6.29}$$

Figs. 6.14 shows the general form of the resulting profile.

As with the non-uniform doping treatment and short channel effect, we would like to express this as an effective change in the body factor γ_s.

$$\gamma_{eff} = \gamma_s\sqrt{1 + \Delta_{RSCE}} \tag{6.30}$$

where

$$\Delta_{RSCE} = 2\frac{L_{rsce}}{L}\left(\frac{N_{rsce}(y)}{N_{ch}} - 1\right)\left(1 - \exp\left[-\frac{L}{L_{rsce}}\right]\right) \tag{6.31}$$

Figs. 6.15 shows how the new RSCE model can be used to match real threshold voltage-channel lengths plots. The distinctive RSCE hump is visible at around 0.1-0.25μm. The STAG3 model does a good job of fitting the experimentally extracted data. SCE fitting parameters were also included to ensure that the simulated threshold voltage drops beyond a certain point.

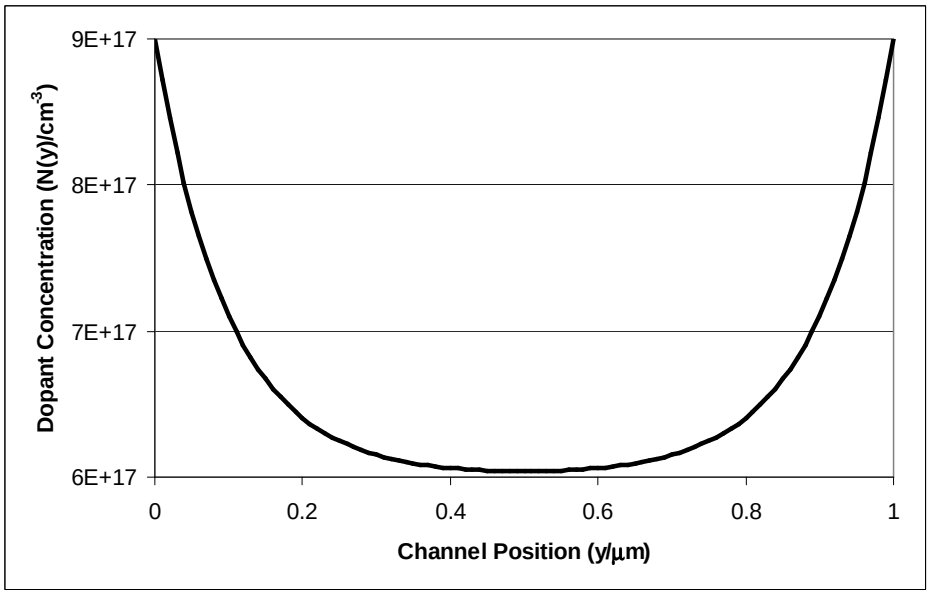

Figure 6.14: *Lateral doping profile from applying Equation (6.31), with dopant pile-up at the source and drain.*

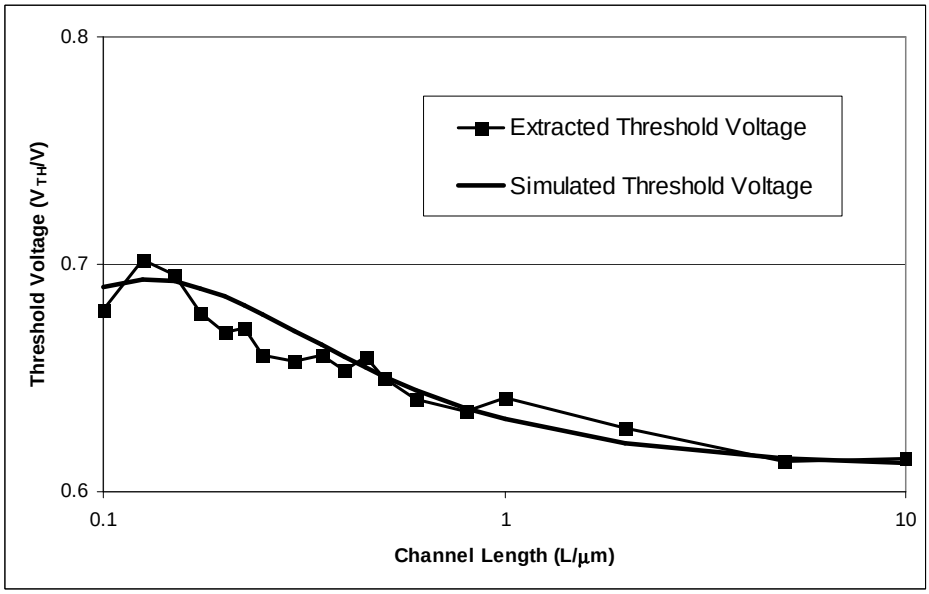

Figure 6.15: *Measured and simulated V_{TH} vs L plots, showing the fitting capability of the new STAG3 RSCE model.*

6.7 Threshold Voltage Extraction Method

Many methods have been proposed to extract the threshold voltage of a MOSFET experimentally [6, 25–28]. Irrespective of the extraction technique used, the measured threshold voltage, V_{Tex}, is conventionally set equal to V_{TH}, the threshold voltage corresponding to the classical criterion for strong inversion [6]

$$V_{TH} = V_{FB} + \psi_{si} + \gamma_s \sqrt{\psi_{si} + V_{SB}} \tag{6.32}$$

where V_{FB} is the flat band voltage, γ_s is the body factor, V_{SB} is the source-body voltage, and ψ_{si} is the surface potential at the onset of strong inversion, usually defined as

$$\psi_{si} = 2\phi_F \tag{6.33}$$

where ϕ_F is the Fermi potential.

It is generally recognised that V_{Tex} will not correspond exactly to V_{TH}; the degree to which V_{Tex} and V_{TH} differ will depend partly on the extraction technique used to obtain V_{Tex}. Compact models based on device surface potential avoid such ambiguity by using the flat band voltage, which has a clear physical definition, to determine the level of inversion in a device. However, in practise the flat band voltage is difficult to extract experimentally, and a common alternative is to extract a value for the threshold voltage and then obtain the flat band voltage using (6.32). Using an unmodified value of V_{Tex} in this way will often result in quite large deviations from the correct flat band voltage, which in turn leads to poor characteristic matching.

We can see the result of such an approach in Figure 6.16. Values for the threshold voltage have been extracted from measured data using three different extraction techniques: the well-known linear extrapolation technique [6, 25, 26], a modified extrapolation technique which employs the device transconductance [27], and the transconductance change technique [28]. Simulations were performed in STAG, using the standard relation (6.32), and compared with the experimental results. It can be seen that the size of the error depends on which extraction technique is being used, but that none of these techniques give a very close match to the measured data.

It has been proposed that a more accurate correspondance between the threshold voltage and the flat band voltage can be obtained by subtracting several ϕ_t from V_{Tex} [25]. However, the main objection to this approach is that it lumps all the errors into a single correction factor, so that it is impossible to separate out different contributions to the threshold shift between V_{Tex} and V_{TH}.

We will now examine how the flat band voltage can be more closely related to the measured threshold voltage by means of a simple physically-based expression. When evaluating our new threshold relation, we will be using the linear extrapolation technique as a reference method, since it is a well-known and widely-used technique [6, 25, 26]. However, before we look at how the threshold model might be improved, let us first examine the existing relation, and the physical assumptions underlying its derivation.

6.7.1 Standard Threshold Voltage Model

Let us consider the case of a body-tied NMOS SOI MOSFET ($V_{SB} = 0$), to which a gate bias V_{GfB} is applied. A small measurement drain voltage V_{DS} is applied to the device.

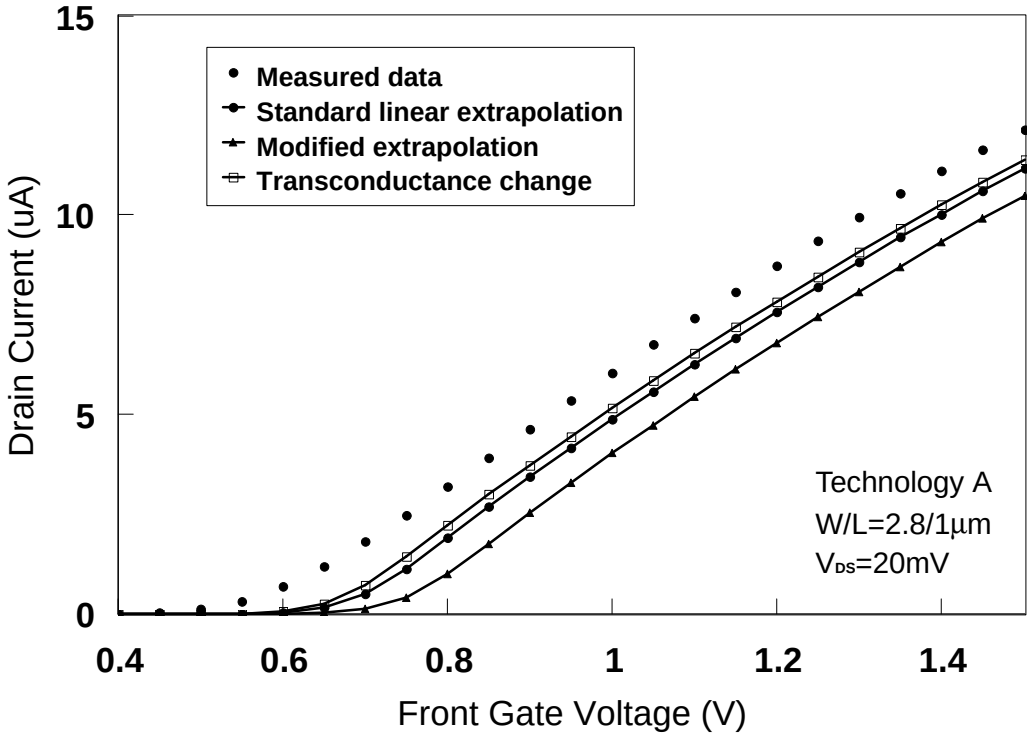

Figure 6.16: *Drain current versus front gate voltage, simulated using unmodified threshold voltages from three different extraction methods, and compared with measured data.*

If we apply Gauss' law and the laws of charge and potential balance across the front gate oxide of the device, we get the following standard result [6, 25]

$$V_{GfB} = V_{FB}^{f} + \psi_s(y) - \frac{q_{tot}(y)}{C_{of}} \tag{6.34}$$

where V_{FB}^{f} is the front gate flat band voltage, C_{of} is the front gate capacitance per unit area, $\psi_s(y)$ is the surface potential, and $q_{tot}(y)$ is the total charge density in the silicon.

Note that because we are applying a voltage between drain and source, the surface potential ψ_s will be a function of the channel position y, as will q_{tot}, since this is itself a function of ψ_s. We can simplify the treatment by assuming the V_{DS} is negligibly small, thus removing the dependency on channel position.

$$V_{GfB} = V_{FB}^{f} + \psi_s - \frac{q_{tot}}{C_{of}} \tag{6.35}$$

The total charge density q_{tot} consists of body and channel charge components q_b and q_c.

$$q_{tot} = q_b + q_c \tag{6.36}$$

The well-known expression for the body charge is used to relate it to the surface potential.

$$q_b = -C_{of}\gamma_s\sqrt{\psi_s} \tag{6.37}$$

Substitute (6.36) and (6.37) into (6.35).

$$V_{GfB} = V_{FB}^f + \psi_s + \gamma_s\sqrt{\psi_s} - \frac{q_c}{C_{of}} \tag{6.38}$$

The threshold voltage V_{TH} is taken to be the value of V_{GfB} at which the channel enters the strong inversion regime. At this point, ψ_s is equal to ψ_{si}, the surface potential at the onset of strong inversion. It is assumed that at the classical onset of strong inversion, the gate voltage is still insufficient to generate a significant inversion layer, and so we consider the inversion charge density q_c to be negligible with respect to the body charge density q_b.

$$q_c \approx 0 \tag{6.39}$$

We can therefore neglect the last term of (6.38). This gives us a general expression relating V_{FB}^f to V_{TH}.

$$V_{FB}^f = V_{TH} - \psi_{si} - \gamma_s\sqrt{\psi_{si}} \tag{6.40}$$

It now only remains to define the value of ψ_{si}, which under the standard convention is taken to be equal to $2\phi_F$. Substituting this into (6.40) gives

$$V_{FB}^f = V_{TH} - 2\phi_F - \gamma_s\sqrt{2\phi_F} \tag{6.41}$$

This standard relation is the one used in STAG2 to obtain the flat band voltage from the value of the threshold voltage used in the parameter set.

6.7.2 Qualitative Features of New Threshold Voltage Model

In order to improve on the existing model, it is necessary to develop a more precise relationship between an actual extracted value of threshold voltage, V_{Tex}, and the flat band voltage, V_{FB}. To achieve this, several adjustments must be made to the standard treatment. These can be summarised in the following three statements.

1. Since a threshold voltage extraction is made using a non-zero drain voltage, the new treatment must account for the influence of a finite drain voltage on the measured threshold voltage. We will use V_{Dex} to designate the drain voltage at which V_{Tex} is extracted.

2. The new treatment must also account for the fact that the value of the extracted threshold voltage usually corresponds to a surface potential which has a value greater than $2\phi_F$. This is because most extraction techniques extrapolate from the strong inversion region, which in reality begins at a higher surface potential than the classical onset point of $2\phi_F$.

We will introduce an empirical parameter, δ_0, which expresses the additional surface potential as a fraction of ϕ_F. The optimal value of δ_0 will vary between different extraction techniques and process technologies; as a result, analytic derivation of δ_0 is not possible. We shall show in this study how it is possible to minimise the range of values which δ_0 can take, by using the new model to account for the different contributions to the threshold shift. In this way, empirical optimisation can be kept to an absolute minimum.

3. Finally, when deriving the standard relation given in (6.41), it is assumed that the inversion charge is negligible compared with the body charge. This is valid when the surface potential is equal to $2\phi_F$, but beyond this point, the inversion charge will rapidly become comparable in magnitude to the body charge. It is therefore not valid to discount the influence of this charge component, especially for larger values of δ_0.

6.7.3 Derivation of New Threshold Voltage Model

As for the standard treatment, we will derive the new relation for the case of a body-tied NMOS SOI MOSFET ($V_{SB} = 0$). A small measurement drain voltage, V_{Dex}, is applied to the device. We once more take (6.34) as our starting point

$$V_{GfB} = V_{FB}^{f} + \psi_s(y) - \frac{q_{tot}(y)}{C_{of}} \tag{6.42}$$

Let us now consider the new form of ψ_{si}, the strong inversion surface potential corresponding to V_{Tex}. Statements 1 and 2 in Section 6.7.2 outline the two primary mechanisms by which $\psi_{si}(V_{Tex})$ and $\psi_{si}(V_{TH})$ will differ. As was the case in the standard treatment, the presence of a drain bias makes the surface potential a function of channel position. We now simplify the treatment, not by neglecting the drain voltage as in the standard treatment, but by averaging V_{Dex} over the entire channel. To account for this average shift in surface potential due to V_{Dex}, we recall that in strong inversion, the drain voltage can be added to the gate induced surface potential [25]. This yields the following expression for ψ_{si}

$$\psi_{si} = (2 + \delta_0)\phi_F + 0.5V_{Dex} \tag{6.43}$$

The surface potential corresponding to V_{Tex} is a constant value, which is necessary since V_{Tex} is itself independent of channel position. The next step is to derive an expression for q_{tot} which is also position independent. Because we cannot neglect the inversion charge (Statement 3 in Section 6.7.2), it is insufficient to simply approximate q_{tot} using the expression for the body charge density, as is done in the standard treatment. Instead, we solve the 1-D form of Poisson's equation in the vertical direction [25], and by again averaging the drain voltage over the channel, we obtain the following result

$$q_{tot} = -\gamma_s C_{of} \sqrt{\psi_{si} + \phi_t \exp\left(\frac{\psi_{si} - 2\phi_F - 0.5V_{Dex}}{\phi_t}\right)} \tag{6.44}$$

Under the new strong inversion condition, (6.42) can be written as

$$V_{Tex} = V_{FB}^{f} + \psi_{si} - \frac{q_{tot}}{C_{of}} \tag{6.45}$$

Substituting (6.43) and (6.44) into (6.45) and rearranging gives us the final result

$$V_{FB}^{f} = V_{Tex} - \psi_{si} - \gamma_s \sqrt{\psi_{si} + \phi_t \exp\left(\frac{\delta_0 \phi_F}{\phi_t}\right)} \tag{6.46}$$

where ψ_{si} is defined in (6.43). For a particular measured device, with a set value of V_{FB}^{f}, we would therefore expect V_{Tex} to increase as V_{Dex} is increased.

For PMOS devices, the sign of the following parameters will be reversed: V_{Tex}, V_{Dex}, ϕ_F, and ψ_{si}. The sign of V_{FB}^{f} will depend on the gate type, the convention being given in [6]. The full PMOS expression is therefore

$$V_{FB}^{f} = V_{Tex} - \psi_{si} + \gamma_s \sqrt{\phi_t \exp\left(\frac{-\delta_0 \phi_F}{\phi_t}\right) - \psi_{si}} \tag{6.47}$$

where

$$\psi_{si} = (2 + \delta_0)\phi_F + 0.5 V_{Dex} \tag{6.48}$$

6.7.4 Evaluation of New Model

Before we test the accuracy of the new expression, one more issue needs to be considered. Although we have accurately accounted for the effect of drain voltage during extraction in an entirely physical manner, there is still an empirical aspect to the new treatment, which is embodied in the parameter δ_0. This parameter can be expected to vary somewhat between different extraction techniques, and also between different technologies.

To reduce the uncertainty associated with this parameter, we can impose a range of values for δ_0 for a given extraction procedure. This was done for the linear extrapolation technique, by testing the new expression against a number of different process technologies, between 0.8μm and 0.25μm. It was found that in all cases, the optimum value of δ_0 required to match the simulated transconductance to the measured data ranged from 0.05 to 0.1, for both NMOS and PMOS devices. We therefore set a default value for δ_0 to 0.05 for this particular extraction method; this provided an good degree of fitting in all the tested technologies. This choice, combined with the factoring out of the drain voltage contribution, allows us to improve the accuracy of our threshold value significantly, without resorting to any optimisation.

Such an approach cannot be used in the standard treatment, since all the contributions to the threshold uncertainty are lumped together. Under the standard regime, therefore, it

would not be possible to distinguish between errors attributable to the selected extraction technique and those resulting from the applied drain bias.

Using the default value for δ_0 of 0.05, and without further optimisation, we can compare the new expression with the standard one; Figures 6.17 and 6.18 make the comparison for two different foundry technologies, which we shall refer to as Technologies A and B (see Appendix A). In each case, a major improvement is seen. Further improvement can of course be obtained with optimisation, but it must be stressed that any optimisation of δ_0 can then be made within a much tighter range of values compared with the standard optimisation of V_{TH} alone. In some cases, the degree of accuracy obtained using the default value will be sufficient. There will of course be other extraction techniques for which the range of δ_0 is higher or lower than it is for the linear extrapolation technique. However, once a range is established for a given technique and a minimum default value deduced, a similar level of accuracy can be obtained, with optimisation only being required for fine-tuning.

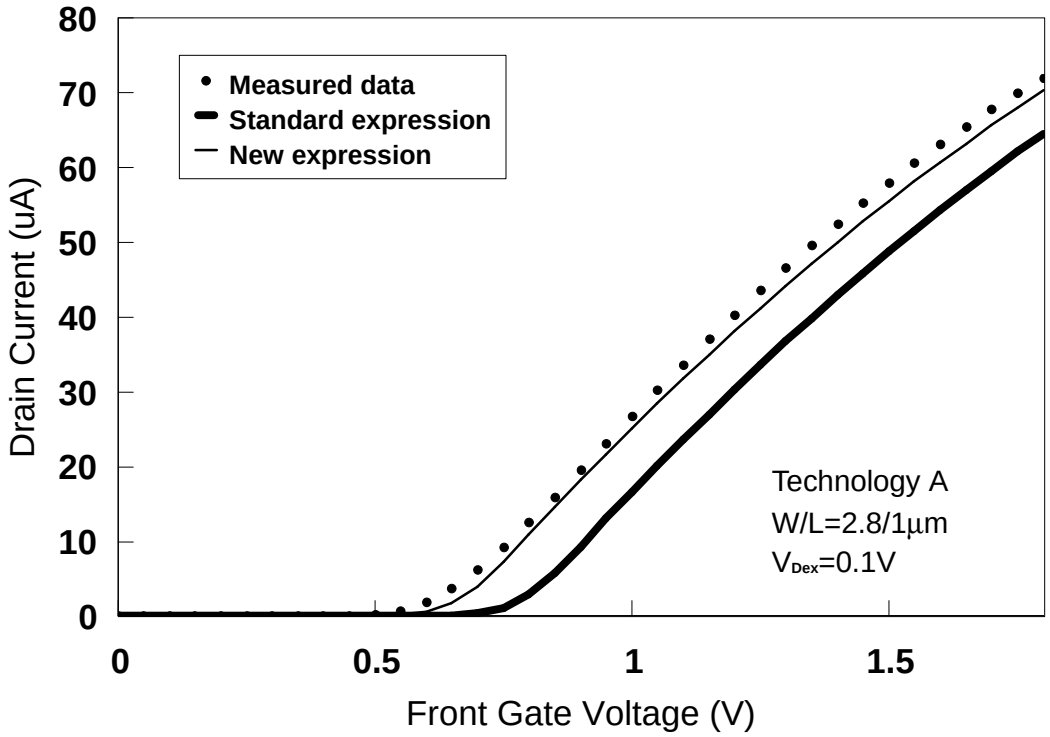

Figure 6.17: *Comparison of measured data with simulation results using new and standard threshold voltage expressions (Technology A). Default value for δ_0 of 0.05 is used.*

That the model does succeed in accounting for the effect of the measurement drain voltage is demonstrated in Figure 6.19. For each of the experimental curves plotted on the figure, obtained using different values of V_{Dex}, a value was extracted for V_{Tex}. Each pair of V_{Dex} and V_{Tex} values correspond to a point on the standard extrapolation line plotted in

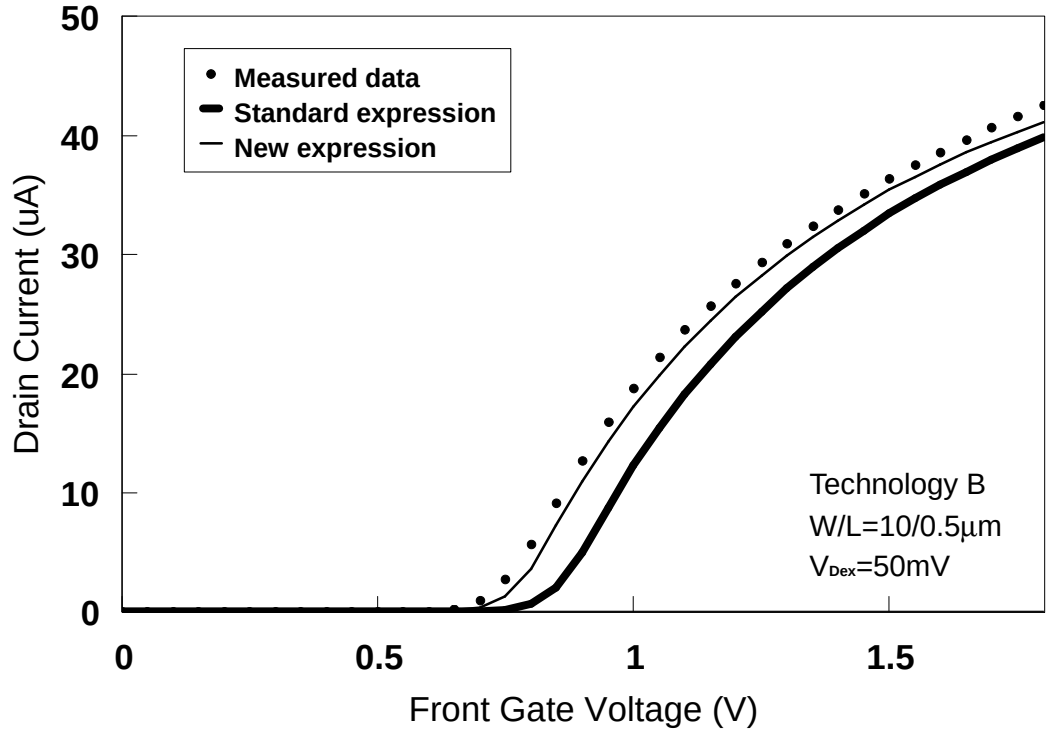

Figure 6.18: *Comparison of measured data with simulation results using new and standard threshold voltage expressions (Technology B). Default value for δ_0 of 0.05 is used.*

Figure 6.20. These pairs of values were then used in simulations to reproduce the experimental curves. This approach has been taken in order to demonstrate that the correct flat band voltage is obtained regardless of the selected value of V_{Dex}. A single value of $\delta_0 = 0.08$, optimised for this particular process and extraction technique, was used for all the simulations. Once δ_0 is set, the variation of V_{Tex} with V_{Dex} is then automatically calculated by the new model equations, resulting in close matching with the experimental data.

We can also examine the way in which V_{Tex} is affected by the drain voltage V_{Dex}. As mentioned in Section 6.7.3, we would expect the extracted threshold voltage to increase with increasing V_{Dex}, provided V_{Dex} is small. This is confirmed experimentally in Figure 6.20,which uses the same three extraction techniques as were used in Figure 6.16. Both the standard and modified extrapolation results indicate a linear relationship, which is reasonable at such low drain voltages. There is slightly more noise associated with the modified extrapolation method, since it involves taking a first derivative of the drain current. The error has been kept small by using small voltage steps and long integration times when obtaining the data used in Figure 6.20. However, if these precautions are not taken, the presence of the first derivative in the extraction equations will lead to less accurate estimates for threshold values. The transconductance change method relies on

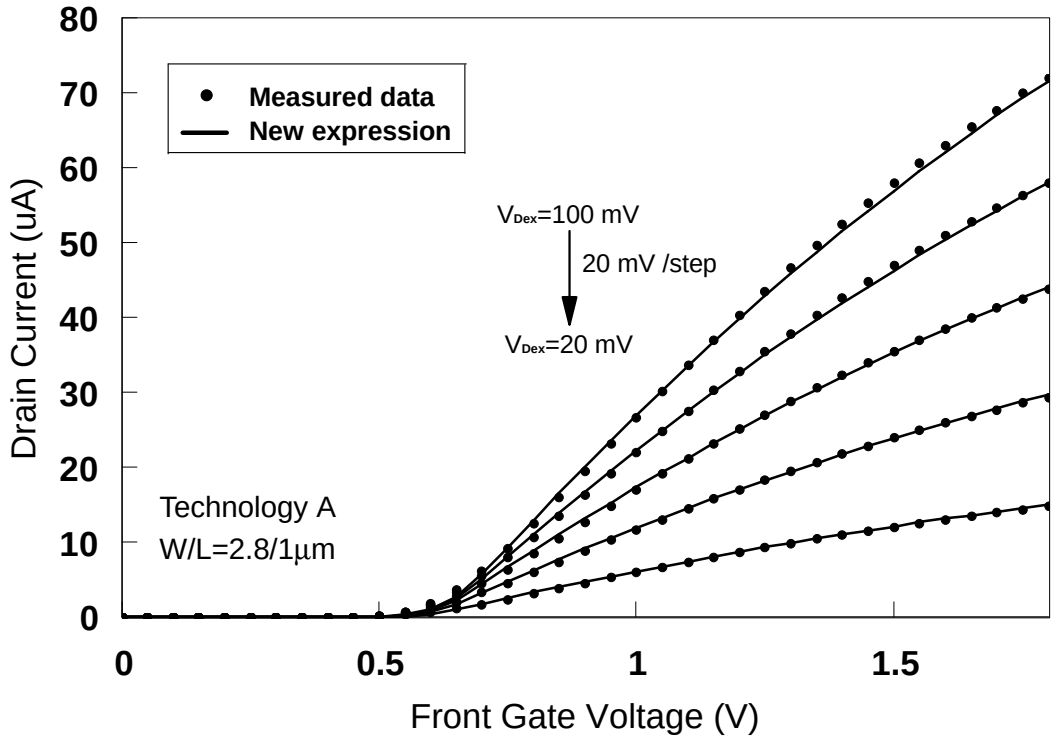

Figure 6.19: *Comparison of measured data with simulation results using the new threshold relation, showing accurate relation between extracted threshold voltage and applied drain voltage.*

the second derivative of the drain current, giving it a larger associated error. As a result, the dependence of V_{Tex} on V_{Dex} is less obvious when using this technique.

The results shown in Figure 6.20 have implications for the selection of a suitable threshold voltage extraction technique. Techniques which require the use of first or higher derivatives of the drain current will possess a larger inherent measurement error compared with those that use the drain current directly. While the new model is able to compensate for systematic errors relating to process technology and extraction method, it cannot account for the random errors inherent to a particular technique. When employing the new threshold relation, it is not necessary to choose a technique with a smaller associated systematic error, which translates to a smaller value of δ_0. Instead the desirability of using a technique with a low random error becomes the guiding criterion when selecting an extraction method. On this basis, the linear extrapolation technique has been shown to work well in conjunction with the new threshold model.

Both the standard and new threshold treatments require that only small drain voltages be used when performing threshold extraction measurements, with an upper limit of 0.1V applying in both cases. This is necessary in the standard treatment in order to justify neglecting the effect of drain bias on the device surface potential. One of the advantages

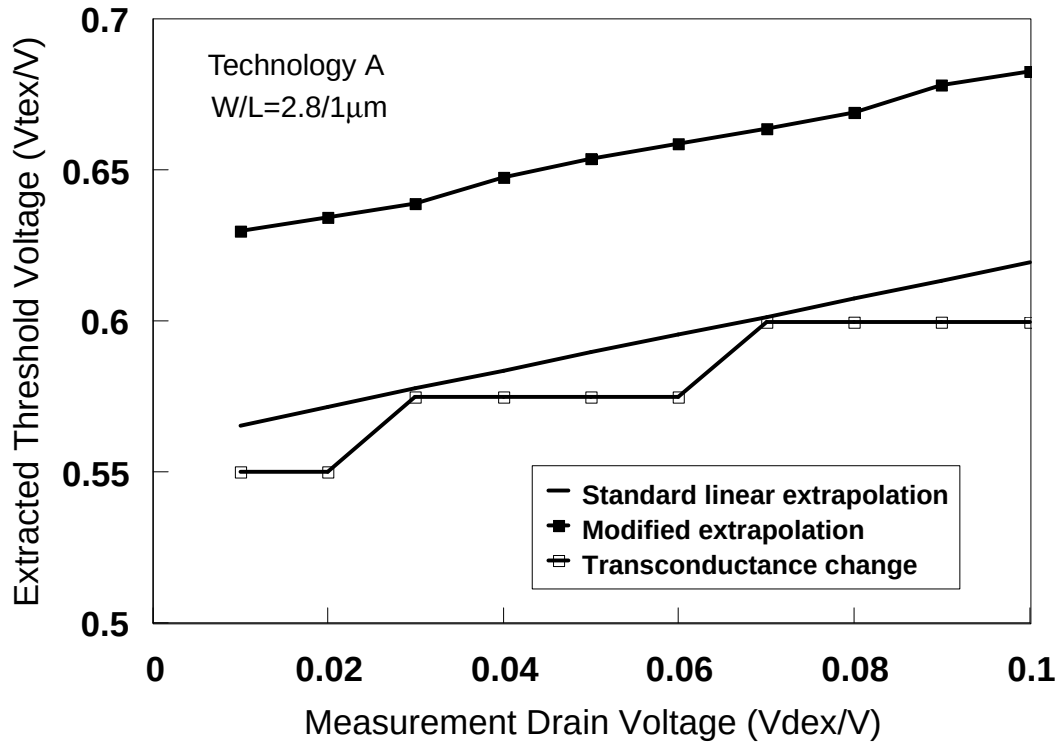

Figure 6.20: *Variation of measured extracted threshold voltage V_{Tex} with drain voltage V_{Dex}, up to V_{Dex}=0.1V. Three different extraction methods were used.*

of the new model is that it recognises that even these small drain voltages can influence the extracted threshold value to a significant degree. However, a small value of V_{Dex} must still be used, because extracted threshold voltages typically correspond to a surface potential only slightly greater than $2\phi_F$. Increasing the drain voltage above this limit may cause the drain end of the channel to move from strong inversion to saturation. We have assumed strong inversion throughout the channel in order to obtain (2.45), because adding the averaged channel voltage to the gate induced surface potential becomes progressively less valid as more of the channel leaves strong inversion [25]. Furthermore, the treatment presented here does not account for drain induced barrier lowering (DIBL) or any of the other effects described in this chapter. At higher drain biases, the DIBL effect will become dominant, and threshold voltage will then decrease with increasing drain voltage.

References

[1] P. Habas, and S. Selberherr, "On The Effect Of Non-Degenerate Doping Of Polysilicon Gate In Thin Oxide MOS-Devices - Analytical Modeling", *Solid-State Electronics*, vol. 33, no. 12, pp. 1539–1544, 1990.

[2] N D Arora, R Rios, and C L Huang, "Modeling the polsilicon depletion effect and

its impact on submicrometer CMOS circut performance", *IEEE Transactions on Electron Devices*, vol. 42, no. 5, pp. 935–943, May 1995.

[3] G Gildenblat, T L Chen, and P Bendix, "Analytical approximation for perturbation of MOSFET surface potential by polysilicon depletion layers", *Electronics Letters*, vol. 35, no. 22, pp. 1974–1976, October 1999.

[4] Jean-Michel Sallese, Matthias Bucher, and Christophe Lallement, "Improved analytical modeling of polysilicon depletion in MOSFETs for circuit simulation", *Solid-State Electronics*, vol. 44, pp. 905–912, 2000.

[5] Ronald R. Troutman, "Ion-Implanted Threshold Tailoring for Insulated Gate Field-Effect Transistors", *IEEE Transactions on Electron Devices*, vol. ED-24, no. 3, pp. 182–192, Mar. 1977.

[6] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[7] N. D. Arora, "Semi-empirical model for the threshold voltage of a double-implanted MOSFET and its temperature dependence", *Solid-State Electronics*, vol. 30, no. 8, pp. 559–569, Aug. 1987.

[8] C. Lallement, M. Bucher, and C. Enz, "Modelling And Characterization Of Non-Uniform Substrate Doping", *Solid-State Electronics*, vol. 41, no. 12, pp. 1857–1861, Dec. 1997.

[9] G. Gildenblat, H Wang, T-L Chen, X. Gu, and X Cai, "SP: An Advanced Surface-Potential-Based Compact MOSFET Model", *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1394–1406, Sep. 2004.

[10] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. J. Scholten, and D. B. M. Klaassen, "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation", *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1979–1993, Sep. 2006.

[11] Reinhard Menzel, "Approximate closed form solution to the error function", *American Journal of Physics*, vol. 43, no. 4, pp. 366–367, Apr. 1975.

[12] Reinhard Menzel, "Erratum: Approximate closed form solution to the error function", *American Journal of Physics*, vol. 43, no. 10, pp. 923, Oct. 1975.

[13] R. K. Bhaduri, and B. K. Jennings, 'Notes on the error function", *American Journal of Physics*, vol. 44, no. 6, pp. 590-592, Jun. 1976.

[14] Richard P. Kenan, "Comment on "Approximate closed form solution to the error function"", *American Journal of Physics*, vol. 44, no. 6, pp. 592, Jun. 1976.

[15] I. Harold Zimmerman, "Extending Menzel's closed-form solution for the error function", *American Journal of Physics*, vol. 44, no. 6, pp. 592-593, Jun. 1976.

[16] John D. Vedder, "Simple approximations for the error function and its inverse", *American Journal of Physics*, vol. 55, no. 8, pp. 762-763, Aug. 1987.

[17] P. Van Halen, "Accurate analytical approximations for error function and its integral", *Electronics Letters*, vol. 25, no. 9, pp. 561-563, Apr. 1989.

[18] M. Nishida, and H. Onodera, "An anomalous increase of threshold voltages with shortening the channel lengths for deeply boron-implanted n-channel MOSFET's", *IEEE Transactions on Electron Devices*, vol. ED-28, pp. 1101–1103, 1981.

[19] C. Y. Lu, and J. M. Sung, "Reverse short-channel effects on threshold voltage in submicrmeter salicide devices", *IEEE Electron Device Letters*, vol. 10, pp. 446–448, 1989.

[20] D. Tsoulakas, C. Tsamis, D. N. Kouvatsos, P. Revva, and E. Tsoi, "Reduction of the Reverse Short Channel Effect in Thick SOI MOSFET's", *IEEE Electron Device Letters*, vol. 18, no. 3, pp. 90–92, 1997.

[21] B. Szelag, F. Balestra, and G. Ghibaudo, "Comprehensive Analysis of Reverse Short-Channel Effect in Silicon MOSFET's from Low-Temperature Operation", *IEEE Electron Device Letters*, vol. 19, no. 12, pp. 511–513, 1998.

[22] S. Kumashiro, "Effect of the transient enhanced diffusion of boron on the characteristics of sub-0.1μm n-MOSFETS", *Materials Science and Engineering*, vol. B71, pp. 148–154, 1999.

[23] N. D. Arora, and M. S. Sharma, "Modeling the Anomalous Threshold Voltage Behaviour of Submicrometer MOSFET's", *IEEE Electron Device Letters*, vol. 13, no. 2, pp. 92–94, 1992.

[24] Hirokazu Hayashi, Noriyuki Miura, Hirotaka Komatsubara, and Koichi Fukuda, "A Simplified Process Modeling for Reverse Short Channel Effect of Threshold Voltage of MOSFET", *IEICE Transactions on Electronics*, vol. E84, no. 9, pp. 1234–1239, 2001.

[25] Y P Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill Book Company, 1987.

[26] D K Schroder, *Semiconductor Material and Device Characterisation*, John Wiley and Sons, 1990.

[27] G Ghibaudo, "New method for the extraction of MOSFET parameters", *Electronics Letters*, vol. 24, no. 9, pp. 543–545, Apr. 1988.

[28] H S Wong, M H White, T J Krutsick, and R V Booth, "Modelling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFETs", *Solid-State Electronics*, vol. 30, no. 9, pp. 953–968, Sept. 1987.

Chapter 7

Auxiliary Model

7.1 Introduction

The core model accounts for the intrinsic device including high field effects, and has been discussed in detail in Chapters 2-6. In this section the auxiliary model is presented, which handles the additional effects outside of the intrinsic MOSFET device. These include series resistance, channel length modulation, parasitic bipolar effect, and self-heating.

7.2 Extrinsic Parasitics and Floating Body Effects

Accurate modelling of floating body behaviour in DC, transient and small signal simulations is essential for a PD-SOI; in all three modes, the observable behaviour will differ profoundly from familiar bulk characteristics [1–3]. Without a sound physical representation of the underlying phenomena, such accuracy cannot be achieved. Furthermore, since digital designs in particular are frequently operated with floating body devices [4], robustness and convergence performance under these conditions must be considered from the outset.

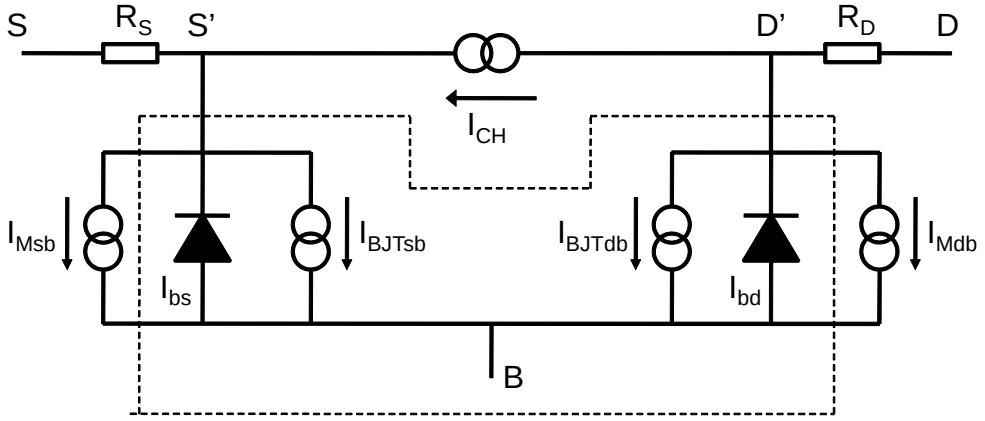

Figure 7.1: *Extrinsic parasitic components included in the STAG3 model.*

The equivalent circuit for the static part of the extrinsic parasitic components is shown in Fig. 7.1 alongside the channel current model. The impact ionisation current is modelled by I_{Mdb} for normal operation and by I_{Msb} for the case $V_{DS} < 0$. This arrangement ensures the symmetry of the model and at all times only one of I_{Mdb} and I_{Msb} will be non-zero. Assuming normal operation, the expression for I_{Mdb} is

$$I_{Mdb} = I_{CH}\left[\frac{\alpha_0}{\beta_0}V_m \cdot \exp\left(-\frac{l_{meff}\cdot\beta_0}{V_m}\right)\right] V_m = V_{DS} - \eta(\psi_{sL\text{sat}} - \psi_{s0}) \tag{7.1}$$

where α_0 and β_0 are the impact ionisation constants, l_{meff} is the effective length of the impact ionisation region, and η is an empirical model parameter used to compensate for errors in the approximation of the lateral field. To improve the fitting capabilities of the model, l_{meff} is defined as

$$l_{meff} = l_m + l_{m1}(V_{DS} - V_{Geff}) + l_{m2}(V_{DS} - V_{Geff})^2 \tag{7.2}$$

where l_m, l_{m1}, and l_{m2} are empirical model parameters, and V_{Geff} is given by

$$V_{Geff} = V_g - V_{SB} - \eta_s 2\phi_F - \gamma_{eff}\sqrt{2\phi_F} \tag{7.3}$$

The body-source and body-drain junction diodes are modelled by I_{bs} and I_{bd} respectively. However, they also form part of the Ebers-Moll model used to account for the lateral parasitic bipolar transistor effect. The expression used for the body-source junction is

$$I_{bs} = \underbrace{I_s\left[\exp\left(\frac{V_{BS}}{\eta_d\cdot\phi_t}\right)-1\right]}_{I_{bs\text{diff}}} + \underbrace{I_{s1}\left[\exp\left(\frac{V_{BS}}{\eta_{d1}\cdot\phi_t}\right)-1\right]}_{I_{bs\text{rec}}} \tag{7.4}$$

The first term models the normal diffusion current mechanism of the diode and the second term accounts for low level recombination. I_s and I_{s1} can be thought of as reverse biased leakage currents and η_d and η_{d1} are the ideality factors with $\eta_d \approx 1$ and $\eta_{d1} \approx 2$. A similar expression is used for the body-drain diode.

For analogue design, the model only needs to indicate if the device is being operated in a regime where bipolar action is dominant. Thus the parasitic bipolar model in STAG acts merely as a warning mechanism and is not a breakdown model involving positive feedback [5]. A modified Ebers-Moll model is used based on the body source and body drain diodes, indicated within the dashed lines in Fig. 7.1. The expressions for the current sources are

$$I_{BJTdb} = \left(\frac{\beta_{BJT\text{eff}}}{\beta_{BJT\text{eff}}+1}\right) I_{bs\text{diff}} \qquad I_{BJTsb} = \left(\frac{\beta_{BJT\text{eff}}}{\beta_{BJT\text{eff}}+1}\right) I_{bd\text{diff}} \tag{7.5}$$

where $\beta_{BJT\text{eff}}$ is the current gain of the BJT. It has been found that $\beta_{BJT\text{eff}}$ exhibits an inverse square dependence on the channel length [6]; therefore STAG uses the following expression

$$\beta_{BJT\text{eff}} = \frac{\beta_{BJT}}{L_{\text{eff}}^2} \tag{7.6}$$

where β_{BJT} is a model parameter.

Note that although a body tie connection from the body region to a convenient reference node (usually the source) will, in principle, eliminate all anomalous behaviour related to the body node, in reality there is always some series resistance. This resistance is technology, layout and indeed bias dependent, and major departures from expected characteristics can arise from its presence [7]. STAG3 does not include this series resistance, therefore if body tie layouts are used, an external resistance should be added based on specific characterisation data.

7.3 Self-Heating Model

The STAG self-heating model is described in detail in [8]. To ensure the consistency and robustness of the model in both static and dynamic analysis modes, the thermal effects have been included directly in the formulation, rather than as a macro model or as a post-processing function.

7.3.1 Basic and Advanced Models

Although the temperature rise in a device is strictly a distributed effect, satisfactory results can be obtained by assuming a uniform average temperature rise. A simple electrical analogue is used in the STAG model involving the basic Joule heating model

$$\Delta T = \Big[I_{CH} \cdot V_{DS} + I_{CH}^2 (R_S + R_D) \Big] \cdot R_T \tag{7.7}$$

where R_T is the device thermal resistance. The temperature rise node is made an external node of the device, allowing both monitoring of each individual device temperature rise and also thermal coupling of the device to neighbouring devices [9].

In cases where multiple time constants have been observed [10], more precise modelling of the thermal behaviour may be needed. STAG allows additional thermal nodes to be introduced, each one possessing an associated thermal resistance and capacitance element. The heating model is implemented as shown in Fig. 7.2 with

$$R_T = R_{T0} + R_{T1} + R_{T2} + R_{T3} + R_{T4} \tag{7.8}$$

Each of these nodes corresponds to an individual thermal time constant, which appears as a pole-zero doublet on a frequency response plot. Under this regime, R_T and C_T effectively become R_{T0} and C_{T0}, with R_{T1-4} and C_{T1-4} then providing the facility to model up to four additional thermal nodes. The existence of each thermal node requires first that its associated resistance be set to a non-zero value, but also that all preceeding thermal resistances be non-zero. For example, setting R_{T2} to zero will not only disable the third thermal node, but will also preclude use of the fourth and fifth thermal nodes as well.

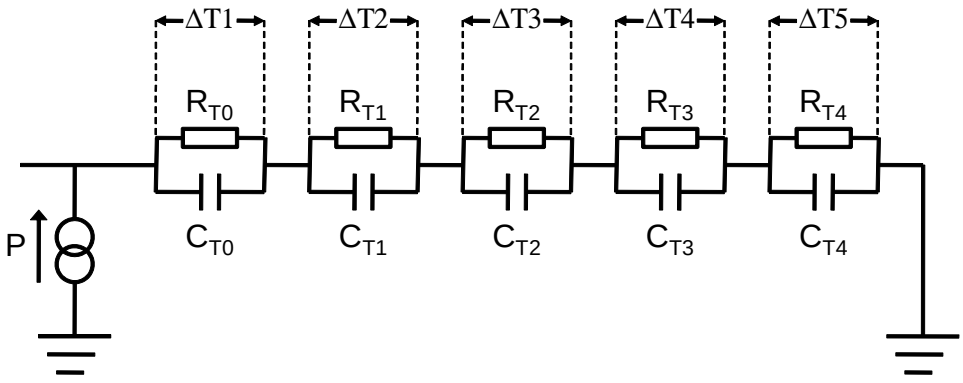

Figure 7.2: *Thermal sub-circuit for STAG3; up to 5 thermal time constants can be modelled.*

Unlike the first-order thermal parameters R_T and C_T, R_{T1-4} and C_{T1-4} automatically default to zero. This arrangement means that the basic self-heating model, which is more appropriate for most circuit design work, is switched on by default, while the more specialised model needs to be activated by the user.

7.3.2 Using the Basic Model

In order to make the self-heating model more useful for modelling and circuit design work, a number of new features have been introduced in STAG3. One feature missing from STAG2 is the facility to calculate default values for thermal resistance R_T and thermal capacitance C_T, in cases where they are not specified. This is useful when performing simulations on circuits containing many transistors, as it can be time-consuming to calculate values for each individual device by hand. Often only an approximate figure is needed for the self-heating parameters, and in these cases a simple first-order model is adequate.

There are several regimes by which R_T and C_T can be specified for any given device in STAG. The following list gives the priorities of these different regimes:

1. By directly specifying R_T and/or C_T, the user overrides all other priorities.

2. If R_T and C_T are not specified, the user can set RTA and CTA, which are the area scaling factors for the thermal resistance and thermal capacitance. RTA and CTA are model parameters; if assigned values in the model netlist, they are used to calculate R_T and C_T for all devices in a circuit using the following relations:

$$C_T = CTA \cdot A_T \tag{7.9}$$

$$R_T = \frac{RTA}{A_T} \tag{7.10}$$

where A_T is the total thermal area of the device defined as being

$$A_T = \left(W + 2 \cdot F_{min} \cdot 10^{-6}\right) \cdot \left(L + 4 \cdot F_{min} \cdot 10^{-6}\right) \tag{7.11}$$

Note that the total thermal area is not simply equal to the front gate area (W multiplied by L), but includes the source/drain contact regions, as well as any overlap around the gate. STAG3 employs an empirical scheme often used by circuit designers to provide an estimate of total device area, based on F_{min}, the minimum feature size of the process technology. The scheme is shown pictorially in Figure 7.3. It is assumed that the front gate is surrounded by a perimeter of silicon material; the width of this border is approximately equal to F_{min}. The drain and source contacts are each assumed to have widths of twice F_{min}. The units of F_{min} are microns (μm); the factors of 10^{-6} are included because W and L are given in metres (m). F_{min} defaults to zero if not defined in the model netlist.

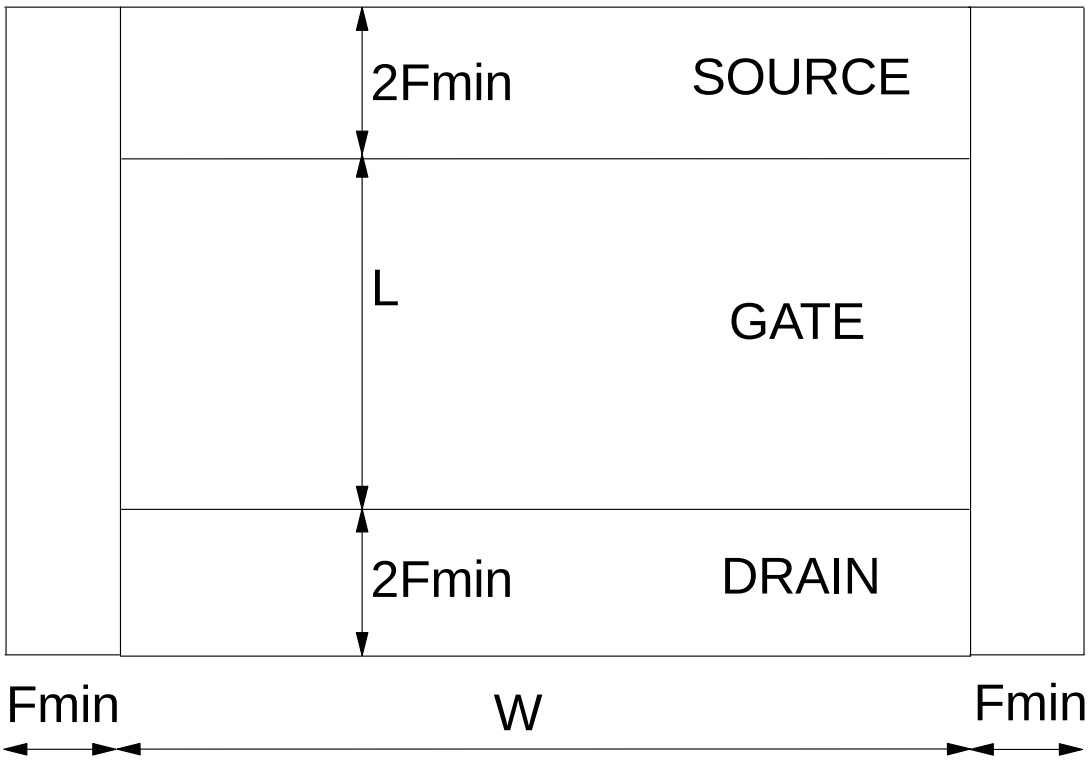

Figure 7.3: *Approximate scheme for calculating the effective thermal area of an SOI device.*

These additional area contributions are important when calculating thermal parameters for devices which approach the minimum feature size in either dimension. Model tests performed on short/narrow channel devices have shown that leaving F_{min} set to zero results in a serious overestimation of R_T, and a corresponding underestimation of C_T. If process layout data is available, then an optimal value of F_{min} may be deduced which may differ from the actual minimum feature size. If, however, no process data is available for a particular technology, it is *strongly recommended* that F_{min} be set equal to the value (in microns) of the process minimum feature width, simply to provide a reasonable order of

magnitude estimate for the self-heating, i.e. when simulating a 0.35μm process, set F_{min} to 0.35. Since F_{min} is also used to calculating default values for the overlap capacitances in STAG3, it is good practice to include a value for this parameter whenever a new netlist is created, irrespective of the self-heating model being used. It is always possible to override any default parameter value calculated from F_{min}.

3. If neither R_T/C_T or RTA/CTA are specified by the user, STAG3 will calculate default values, using a first-order physical model. The relations used are

$$C_T = \rho_{si} \cdot c_{si} \cdot A_T \cdot t_b \tag{7.12}$$

$$R_T = \frac{t_{ob}}{\kappa_{ox} \cdot A_T} \tag{7.13}$$

where ρ_{si} is the density of silicon ($2330 kgm^{-3}$), c_{si} is the heat capacity per unit mass of silicon ($700 JK^{-1}kg^{-1}$), and κ_{ox} is the thermal conductivity of the silicon oxide ($1.4 WK^{-1}m^{-1}$). t_b and t_{ob} are the model parameters specifying the silicon film thickness and back oxide thickness respectively.

It should be understood that the first order self-heating model is switched *on* by default; if R_T and C_T are not specified by the user, STAG3 will calculate them using the above default model.

4. To switch off self-heating for a single SOI device, simply set R_T to zero. As discussed in the previous section, this will eliminate all the thermal nodes from that device, regardless of the settings of the other thermal parameters. To switch off self-heating for all SOI devices, set RTA to zero. Be aware that this will only work for devices for which R_T is not set; because of the order of priorities, any device with a non-zero R_T will ignore the RTA setting.

To switch off dynamic self-heating effects only, for a single device, set C_T to zero, whilst leaving R_T set. To do this for all devices, set CTA to zero, whilst leaving RTA set.

Note that the preceeding options apply to the basic, single node model only. The multiple node model is chiefly intended for detailed modelling of thermal characteristics, and is appropriate when multiple thermal time constants have been extracted experimentally and need to be incorporated into simulations. In such instances, it will usually be more appropriate to use extracted values to specify R_T, R_{T1}, R_{T2} etc, rather than rely on a scaling model. For this reason, no scaling model is provided for R_{T1-4} and C_{T1-4}. The area scaling regime can still be applied to R_T and C_T when the other nodes are set (provided of course that RTA/CTA or the default values are used).

7.3.3 Thermal Effect on Physical Parameters

The STAG model implements dependence of threshold voltage (via the flat-band voltage), mobility and carrier saturation velocity on the local device temperature. Being a physically based model, these three quantities should be sufficient to model the thermal dependence of the intrinsic device [11, 12].

Threshold Voltage

The thermal dependence of the measured device threshold is accounted for by the use of an effective flat-band voltage.

$$V^f_{FB\text{eff}} = V^f_{FB} - \frac{\sigma}{L} \cdot V_{DS} + \chi_{FB} \cdot \Delta T \tag{7.14}$$

where the shift due to DIBL is included for completeness.

Mobility

For a given temperature rise, this physical quantity has by far the most effect on the output characteristics of a device. It has been found that the surface mobility in a MOSFET varies as T^{-k} [13], with k reported between 1.4 to 1.8. In STAG, the mobility thermal dependence is written as

$$\mu_{0\text{th}} \equiv \mu_0(T) = \mu_0(T_{\text{amb}}) \cdot \left(1 + \frac{\Delta T}{T_{\text{amb}}}\right)^{-k} \tag{7.15}$$

k has been left as a model parameter in STAG, but it defaults to 1.5. The thermal dependencies of each scattering mechanism hae already been discussed in Chapter 3.

Carrier Saturation Velocity

In STAG, it is assumed that a saturation velocity has been found at T_{amb}, being denoted by v_{sat0}, and this value is related to the value used in the model by

$$v_{\text{sat}} = \frac{v_{\text{sat0}} \cdot \left[1 + 0.8 \exp\left(\frac{T_{\text{amb}}}{600}\right)\right]}{1 + 0.8 \exp\left(\frac{T_{\text{amb}} + \Delta T}{600}\right)} \tag{7.16}$$

This expression reflects the functional temperature dependency reported in the literature [13].

7.3.4 Thermal Effect on Parasitic Components

The local temperature rise ΔT will affect the parameters of the parasitic components of the device as well as the intrinsic device. Investigations into the impact ionisation current [14, 15] indicate that the only parameter which exhibits a temperature dependence is β_0. This dependence is linear and so the STAG model uses the following equation to incorporate this effect.

$$\beta_M = \beta_0 + \chi_\beta \cdot \Delta T \tag{7.17}$$

where χ_β is a model parameter. β_M is then substituted for β_0 in Equation 7.1.

There are two quantities in the ideal diode equation which exhibit a temperature dependence. The obvious one is $\phi_t = kT/q$, and the other is the reverse saturation current I_s and I_{s1}. It has been found [13] that

$$I_s \sim T^{(3+p/2)} \cdot \exp\left(-\frac{E_g}{kT}\right) \tag{7.18}$$

where p is a constant as specified in [13]. At normal operating temperatures however, the exponential term dominates, and hence the STAG model uses a simple exponential temperature dependence.

7.4 Channel Length Modulation

The effect of channel length modulation (CLM) is modelled using a variant of the Klaasen model [16, 17], suitably modified to avoid numerical difficulties. An alternative linear 'λ' model for CLM [18] can also be used within STAG to facilitate fast parameter extraction at the expense of some accuracy.

The effect of channel length modulation (CLM) on channel current is included in STAG3 in the following way

$$I_{CH} = \left(1 + \frac{l_d}{L_{\text{eff}}}\right) I_{CH\text{sat}} \tag{7.19}$$

where l_d is the length of the saturation region, and L_{eff} is the effective channel length of the device.

Two CLM models are available in STAG, the standard lambda model, and a more advanced model intended to provide a better fit for sub-micron devices. In the lambda model, l_d is determined according to the following equation

$$l_d = \lambda \cdot (V_{DS} - V_{D\text{sat}}) \tag{7.20}$$

where λ is a model parameter, and $V_{D\text{sat}}$ is the saturation voltage, which in STAG is expressed in surface potential form.

$$V_{D\text{sat}} = \psi_{sL\text{sat}} - \psi_{s0} + \phi_t \tag{7.21}$$

In the sub-micron model, l_d is determined as follows

$$l_d = l_x \cdot \ln\left(1 + \frac{V_{DS} - V_{DSlim}}{v_p}\right) \tag{7.22}$$

The quantities l_x and v_p are model parameters. V_{DSlim} is given by

$$V_{DSlim} = V_{DS} \cdot V_{D\text{sat}} \cdot \exp\left(\frac{-\ln(V_{DS}^k + V_{D\text{sat}}^k)}{k}\right) \tag{7.23}$$

In the above equation, k = 2·MEXP, where MEXP is a model parameter which should be set to an integer value. For short-channel devices, MEXP=1 has been found to work well.

Which CLM model is selected depends on which model parameters are present in the netlist. The lambda model has priority, so that if parameters from both models are selected, the lambda model will be used. There are three possible CLM model configurations in STAG3.

1. If λ is set in the model netlist, and MEXP is not set, then the standard lambda model will be used.
2. If λ is set in the model netlist, and MEXP is set to an integer value of at least 1, then the lambda model will be used, but with $V_{D\text{sat}}$ being replaced by V_{DSlim}.
3. If λ is not set, but l_x and v_p are, and MEXP is set to an integer value of at least 1, then the sub-micron model will be used.

7.5 External Source and Drain Series Resistance

STAG3 models the effect of source and drain series resistance using two lumped resistances, R_S and R_D. It is possible to specify these parameters using three regimes, which are described below in order of priority.

1. By directly specifying R_S and R_D, the user overrides all other priorities.

2. The user can define the series resistance in terms of the sheet resistance RSH.

$$R_S = NRS \cdot RSH \tag{7.24}$$

$$R_D = NRD \cdot RSH \tag{7.25}$$

where NRS/NRD are the number of squares associated with the source and drain (these are specified as instance parameters for each device).

3. The user can ensure that series resistances scale automatically with device width by using the source/drain series resistance scaling factors, RSW/RDW. To calculate values for RSW/RDW, we extract R_S and R_D for a single device. We define RSW as the product of the source series resistance (ohms), and the width (microns) of that device. RDW is similarly obtained by multiplying R_D and W for the same device. Once these values are in place, scaling is automatically performed for any device, using the following relation

$$R_S = \frac{RSW}{W \cdot 10^{-6}} \tag{7.26}$$

$$R_D = \frac{RDW}{W \cdot 10^{-6}} \tag{7.27}$$

Thus, by introducing RSW and RDW as STAG3 model parameters, we remove the need to perform manual calculations of R_S and R_D.

References

[1] J Tihanyi and H. Schlotterer, "Properties of EFSI MOS transistors due to the floating substrate and the finite volume", *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 1017–1023, 1975.

[2] A Wei, M J Sherony, and D A Antoniadis, "Transient behavior of the kink effect in partially-depleted SOI MOSFET's", *IEEE Electron Device Letters*, vol. 16, no. 11, pp. 494–496, Nov. 1995.

[3] R Howes and W Redman-White, "A small-signal model for the frequency-dependent drain admittance in floating-substrate MOSFETs", *IEEE Journal of Solid-State Circuits*, vol. 27, no. 8, pp. 1186–1193, Aug. 1992.

[4] D. Allen, A. Aipperspach, D. Cox, N. Phan, and S. Storino, "A 0.20μm 1.8V SOI 550MHz 64b PowerPC microprocessor with Cu interconnects", in *Proceedings ISSC Conference*, 1999, pp. 438–439.

[5] J-P Colinge, *Silicon-On-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers, 1991.

[6] J Chen, F Assaderaghi, P K Ko, and C Hu, "The enhancement of gate-induced-drain-leakage (GIDL) current in short-channel SOI MOSFET and its application in measuring lateral bipolar current gain β", *IEEE Electron Device Letters*, vol. 13, no. 11, pp. 572–574, 1992.

[7] C F Edwards, B M Tenbroek, M S L Lee, W Redman-White, and M J Uren, "The effect of body contact series resistance on SOI CMOS amplifier stages", *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2290–2294, Dec. 1997.

[8] M S L Lee, *Compact Modelling of Partially Depleted Silicon-On-Insulator MOSFETs for Analogue Circuit Simulation*, PhD thesis, University of Southampton, Southampton, UK, Dec. 1997.

[9] B M Tenbroek, M S L Lee, W Redman-White, C F Edwards, R J T Bunyan, and M J Uren, "Measurement and simulation of self-heating in SOI CMOS analogue circuits", in *Proceedings IEEE international SOI conference*, Fish Camp, California, USA, Oct. 1997, pp. 156–157.

[10] A L Caviglia and A A Iliadis, "Linear dynamic self-heating in SOI MOSFETs", *IEEE Electron Device Letters*, vol. 14, no. 3, pp. 133–135, 1993.

[11] B M Tenbroek, M S L Lee, W Redman-White, R J T Bunyan, and M J Uren, "Self-heating effects in SOI MOSFET's and their measurement by small signal conductance

techniques", *IEEE Transactions on Electron Devices*, vol. 43, no. 12, pp. 2240–2248, Dec. 1996.

[12] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[13] S.M. Sze, *Physics of Semiconductor Devices*, Wiley, 2nd edition, 1981.

[14] R Howes, W Redman-White, K G Nichols, P J Mole, M J Robinson, and S Bird, "An SOS MOSFET model based on calculation of the surface potential", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 6, pp. 494–506, Apr. 1994.

[15] R Howes, *Modelling and simulation of Silicon-On-Sapphire MOSFETs for analogue circuit design*, PhD thesis, University of Southampton, Southampton, UK, Aug. 1991.

[16] F M Klaassen, P T J Biermans, and R M D Velghe, "The series resistance of submicron MOSFETs and its effect on their characteristics", in *Proceedings European Solid State Device Research Conference*, 1988, pp. 257–260.

[17] H C de Graaf and F M Klaassen, *Compact Transistor Modelling for Circuit Design*, Computational Microelectronics. Springer-Verlag, Wien, New York, 1990.

[18] Y P Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill Book Company, 1987.

Chapter 8

Charge Model

8.1 Capacitance Modelling in SPICE

SPICE-based compact MOS models generally utilise two types of capacitance model - intrinsic and extrinsic. Intrinsic capacitances are associated with the movement of charge within the silicon device, and are can be derived from the expressions for channel and body charge. Extrinsic capacitances are usually handled outside of the the main charge model, and are used to account for parasitic contributions, such as overlap capacitance and junction depletion capacitance.

8.2 Intrinsic Capacitances

In SPICE models, an intrinsic capacitance C_{ij} is implemented according to the definition

$$C_{ij} = \frac{dQ_i}{dV_j} \tag{8.1}$$

where Q_i is the total charge associated with node i, and V_j is the small-signal voltage at node j, referred to the other nodes which are held at a steady DC bias.

Since we are considering floating-body behaviour, our treatment is only relevant to partially-depleted devices, and we will therefore neglect the influence of the back gate. This leaves four terminal nodes, and therefore sixteen intrinsic capacitances. Of these, four are conventional capacitance elements: C_{ss}, C_{dd}, C_{gfgf}, and C_{bb}. The remaining twelve are transcapacitances, which model the variation of charge at one node with respect to the small-signal voltage at another node. It should be noted that these transcapacitances are non-reciprocal, so that in general

$$C_{ij} \neq C_{ji} \tag{8.2}$$

It has been shown that this condition of non-reciprocity is linked to the issue of charge conservation [1], and that reciprocal models, such as the Meyer model [2], give non-physical results. As a result, the inclusion of non-reciprocal capacitance elements in SPICE-based

models is critical to allow accurate modelling of small-signal effects.

For the modelling of the dynamic charge equations, the approach taken closely follows that used in the SUSOS model [3]; charge non-conservation is avoided by modelling the transient currents as the time derivatives of the nodal charges.

In order to calculate the transcapacitances, we require analytic expressions for the nodal charges. These are obtained by integrating the charge densities q_b and q_c along the length and width of the device. As was explained in Chapter 2, the active length of the device is split into two distinctive regions. The GCA region (i.e. that portion of the channel where the Gradual Channel Approximation is taken to be valid), is the region for which the core and high field models in Chapters 2-6 were developed. The second region is the drain region, where the lateral drain field is at least comparable to the gate, and for which the GCA approximation is not valid. In this second region, it is necessary to introduce empirical models such as the CLM model in order to account for high lateral field effects.

The length of the GCA region is denoted by L', and the drain region by l_d. We have already defined l_d in Section 7.4, using either Equation (7.20) or Equation (7.22), depending on which CLM model we have selected. L' is defined as

$$L' = \frac{L}{\left(1 + \frac{l_d}{L}\right)} \tag{8.3}$$

The above expression is the same as that used in STAG2. In keeping with the convention used in STAG2, the charges associated with the GCA region will be denoted by the suffix '1', while the drain region will use the suffix '2'.

8.2.1 GCA Region Charge

The total channel charge in the GCA region can be expressed as follows:

$$Q_{CH1} = \int_0^{L'} q_c dy \tag{8.4}$$

It is necessary to partition this charge between the course and drain node such that

$$Q_{D1} + Q_{S1} = Q_{CH1} \tag{8.5}$$

Various partitioning schemes have been proposed; we have chosen to use the commonly employed Ward-Dutton treatment [4]. While other more sophisticated schemes are available, it has been shown that this method gives accurate results, provided that the variation of surface potential with respect to position along the channel is correctly modelled [5].

Using the Ward-Dutton charge partitioning scheme yields the following relations for the drain and source charge

$$Q_{D1} = \int_0^{L'} \frac{y}{L'} q_c dy \tag{8.6}$$

$$Q_{S1} = \int_0^{L'} \left(1 - \frac{y}{L'}\right) q_c dy \tag{8.7}$$

Because we have defined the channel charge as a function of the surface potential, we need to find a way to express it as a function of y, the position along the channel. We start with the charge sheet model expression for channel current from Equation (2.1), and neglect the diffusion contribution to give only the drift term

$$I_{CH}(y) \approx -\mu_s W q_c(y) \frac{d\psi_s(y)}{dy} \tag{8.8}$$

We remove the position dependence by integrating from y=0 to y=L', and then dividing by L'

$$I_{CH} = -\mu_s \frac{W}{L'} \int_{\psi s0}^{\psi_{sL}} q_c(\psi_s) d\psi_s \tag{8.9}$$

We now change variables, and substitute Equation (2.19) into Equation (8.9)

$$I_{CH} = -\mu_s \frac{W}{L'} \frac{1}{\alpha C_{of}} \int_{q_s}^{q_d} q_c dq_c \tag{8.10}$$

where q_s is the channel charge density at the source end of the channel

$$q_s = -C_{of}[V_{GT} - \alpha\psi_{s0}] \tag{8.11}$$

and q_d is the channel charge density at the drain end of the channel

$$q_d = -C_{of}[V_{GT} - \alpha\psi_{sL}] \tag{8.12}$$

We can also integrate Equation (8.10) to some point y in the GCA region of the channel

$$I_{CH} = -\mu_s \frac{W}{y} \frac{1}{\alpha C_{of}} \int_{q_s}^{q_c(y)} q_c dq_c \tag{8.13}$$

Assuming that the device is quasi-static, I_{CH} must be the same at every point along the channel. Therefore we can equate Equation (8.10) and Equation (8.13) to give

$$\frac{y}{L'}\left[q_d^2 - q_s^2\right] = q_c^2 - q_s^2 \tag{8.14}$$

Rearranging allows us to formulate an expression for q_c as a function of y

$$q_c = q_s \sqrt{1 + \frac{y}{L'} \left[\frac{q_d^2}{q_s^2} - 1 \right]} \tag{8.15}$$

Substituting Equation (8.15) into Equation (8.4) gives us our final expression for the channel charge

$$Q_{CH1} = WL' \left[-\frac{2}{3} q_s \left(\frac{F^2 + F + 1}{F + 1} \right) \right] \tag{8.16}$$

where

$$F = \frac{q_d}{q_s} \tag{8.17}$$

Similarly, by substituting Equation (8.15) into Equation (8.7), and integrating by parts, gives us our final expression for the drain charge

$$Q_{D1} = WL' \frac{2}{15} q_s \left[\frac{3F^3 + 6F^2 + 4F + 2}{(F + 1)^2} \right] \tag{8.18}$$

The total body charge in the GCA region can be expressed as follows:

$$Q_{B1} = W \int_0^{L'} q_b dy \tag{8.19}$$

As before, we have the problem that q_b in Equation (2.18) is expressed in terms of ψ_s. We use a similar procedure to that used for Q_{CH1} in order to change the variable to y. Equation (8.19) then yields

$$Q_{B1} = WL' C_{of} \gamma_s \left\{ \frac{\delta_s}{\alpha} \left[\frac{2}{3} \frac{q_s}{C_{of}} \left(\frac{F^2 + F + 1}{F + 1} \right) + V_{GT} \right] - \left(\sqrt{\psi_{st0}} - \delta_s \psi_{st0} \right) \right\} \tag{8.20}$$

8.2.2 Drain Region Charge

The treatment for the drain region charge is very simple; since the carriers are velocity saturated, the channel charge density is assumed to be uniform in the drain region. Since we already know q_d, the channel charge density at the drain end of the GCA region, we use this quantity. The simple STAG2 model has been retained, whereby the drain channel charge is partitioned equally between the source and drain nodes.

$$Q_{S2} = Q_{D2} = \frac{Q_{CH2}}{2} = \frac{W l_d q_d}{2} \tag{8.21}$$

The drain body charge is a more complicted issue, since depletion charge would be shared with the drain p-n junction depletion region, which is handled by the extrinsic charge model. STAG2 sets Q_{B2} to zero, and relies on setting the junction capacitance charge to compensate. As we shall see in the detailed analysis given in the next section, this approach can lead to problems. Here we shall simply present the expression used in STAG3 for the

drain body charge density, with the derivation to follow.

$$Q_{B2} = \gamma_s C_{of} W F_{CS} l_d \left[\sqrt{\psi_{st0}} + \delta_s \left(\psi_{sL} - \psi_{st0} \right) \right] \tag{8.22}$$

Since we do not take account of the back gate charge, we can now compute the front gate charge using the law of charge conservation.

$$Q_G = -\left(Q_{CH1} + Q_{CH2} + Q_{B1} + Q_{B2} \right) \tag{8.23}$$

8.2.3 Drain Region Body Charge Model

A number of floating body effects have been observed in Partially-Depleted SOI devices which lack a body contact, or those in which the contact is ineffective [6]. Of these, perhaps the best understood is the kink effect, which has received a great deal of attention previously [7–9]. In the small-signal regime, the variation of output conductance with frequency is very pronounced in the kink region of a floating body device, and if compact models are to accurately reflect the behaviour of floating body devices, they need to be able to account for this effect.

However, most analogue circuits require that devices be biased in the saturation region, below the onset of kink. This ensures that devices are not subjected to large variations in their output conductance whilst biased in saturation [10]. Hence the small-signal effects which occur in saturation *below* the onset of kink will have a large influence on the characteristics of an analogue circuit designed in a floating body MOS technology.

Despite being of interest to analogue designers, the small-signal behaviour of floating body devices in the below-kink saturation region has received relatively little coverage. It has been shown [9] that when measuring saturation output conductance, g_{out}, as a function of frequency, a pole/zero pair is seen, as shown by the experimental data in Figure 8.1. This feature appears due to variations in the small-signal body potential with frequency, which modulates the output conductance via the body transconductance. This behaviour is not observed in bulk MOS devices or body-tied SOI where there is no floating body node; instead the conductance response is essentially flat over this range. In modern SOI devices, the pole/zero pair occurs in the 1-10Hz frequency range, and thus is difficult to observe directly unless equipment is used which is capable of taking measurements at such low frequencies. In order that the frequency response be clearly seen, the measurements shown in Figure 8.1 are for a floating-body Silicon-on-Sapphire (SOS) device, possessing body junction leakage currents than are higher than modern SOI processes; hence the pole is shifted to a higher frequency [9]. Any compact model intended for analogue design in SOI must be able to account for this variation in the saturation output conductance. Simple qualitative agreement is not usually sufficient, since the characteristics of many circuits (i.e. amplifier gain) will be directly determined by the magnitude of g_{out}.

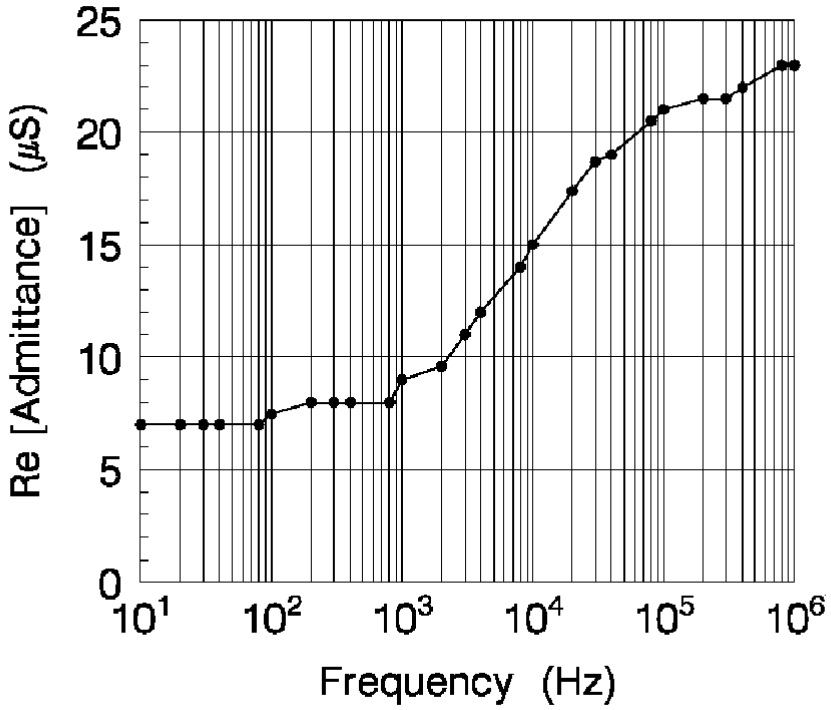

Figure 8.1: *Measured small-signal output conductance versus frequency for a floating body SOS device biased in saturation, below the onset of kink (W/L=20/3μm, V_{GS}=2V, V_{DS}=2V). Taken from Reference [9].*

Floating Body Small-Signal Equivalent Circuit

Although the equivalent circuit representation used in [9] is sufficient to qualitatively predict the AC behaviour of floating body devices, for frequencies up to about 1-10 MHz, it considers only the use of reciprocal capacitance elements and so does not reflect the use of transcapacitances in SPICE models. In this study we will be using a modified circuit, shown in Figure 8.2, which does include non-reciprocal elements. This approach allows us to derive a mathematical relation between the SPICE intrinsic capacitance model and the simulated output conductance response.

The basic circuit configuration is the same one used in [9]. The source and gate terminals are tied to AC ground, and taken as the small-signal reference, while an AC voltage is applied to the drain. DC biases are applied to the gate and drain such that the device is operating in saturation. We have neglected the presence of source and drain series resistance; although they can be included, they complicate the mathematical analysis, and do not contribute to our understanding of the key issues. This simplifies the treatment considerably, as it eliminates the effect of the front gate transconductance, and means that only two intrinsic capacitances appear in the equivalent circuit. Since we are interested in the below-kink saturation region, we will also neglect the influence of impact ionisation and parasitic BJT action.

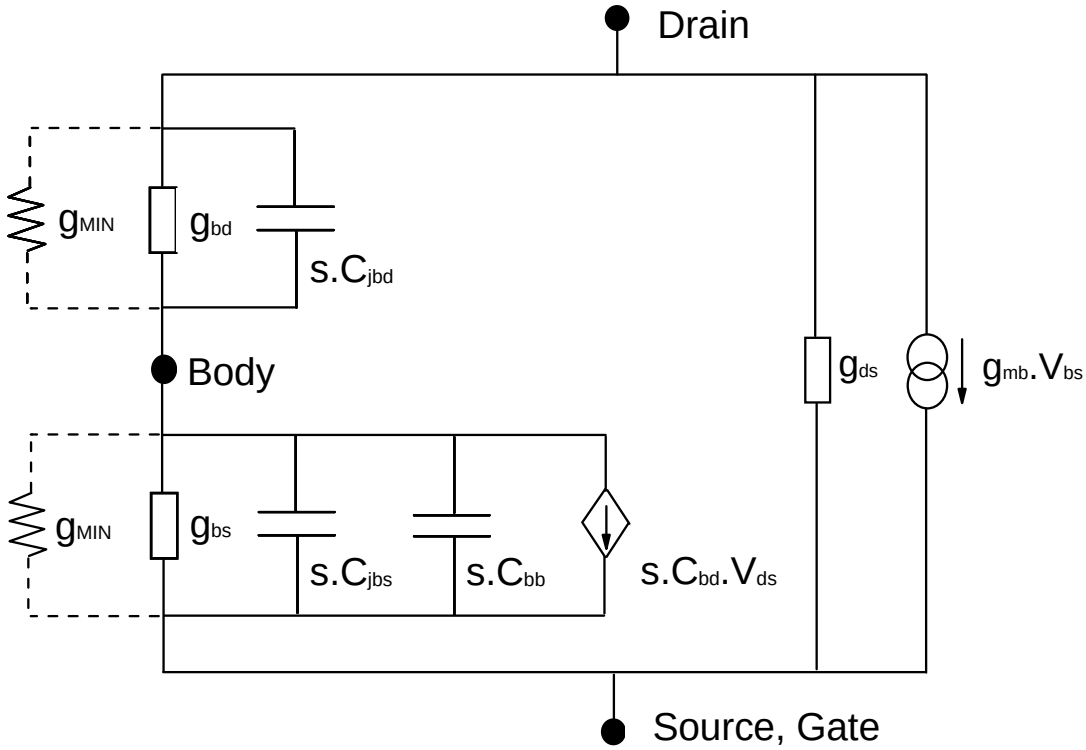

Figure 8.2: *Equivalent small-signal circuit for a floating body SOI MOSFET with the source and front gate tied to AC ground.*

The circuit elements in Figure 8.2 are as follows: g_{ds} is the internal drain-source conductance, g_{mb} is the body-source transconductance, g_{bs} and g_{bd} are the body diode conductances, and g_{MIN} is the minimum nodal conductance employed by SPICE to improve convergence behaviour. The influence of g_{MIN} on floating body simulations will be discussed in Section 8.2.3; for the moment, we will assume that it is set to a negligible value. There are two intrinsic capacitances C_{bb} and C_{bd}, and two extrinsic junction capacitances, C_{jbs} and C_{jbd}. Note that (8.1), the transcapacitance definition as implemented in SPICE, yields the opposite sign to the convention commonly used in literature [11] when applied to non-reciprocal capacitances. In Figure 8.2 therefore, we have reversed the direction of the current source associated with C_{bd}.

Derivation of Small-Signal Drain Admittance

The general form for the external drain-source admittance of a floating body SOI MOSFET is given by (3) in [9] as

$$y_{out} = g_{ds} + \frac{1}{Z_{bs} + Z_{bd}} + g_{mb} \cdot \frac{Z_{bs}}{Z_{bs} + Z_{bd}} \tag{8.24}$$

where Z_{bs} and Z_{bd} are the body-source and body-drain network impedances. This ex-

pression only holds if there are no transconductive or transcapacitive elements present in either body network.

Three terms contribute to the total drain admittance in Equation (8.24). The first term is the internal drain-source conductance g_{ds}. The second term is the contribution from the direct path through the body network. The third term is the contribution of the body-source transconductance g_{mb}, modulated by the body-source bias, which in turn is governed by the potential divider ratio of the body network impedances Z_{bs} and Z_{bd}. We can neglect the second term, since it is typically much smaller than g_{ds} and g_{mb} [9]. This approximation is quite valid when applied to frequencies in the range depicted in Figure 8.1.

We will now obtain an expression for the drain admittance, y_{out} in our equivalent circuit in Figure 8.2. Let us write an expression for the drain-source current

$$I_{ds} = g_{ds} \cdot V_{ds} + g_{mb} \cdot V_{bs} + g_{bs} \cdot V_{bs} + s \cdot (C_{bb} \cdot V_{bs} + C_{bd} \cdot V_{ds} + C_{jbs} \cdot V_{bs}) \tag{8.25}$$

We can simplify Equation (8.25) by neglecting the current component passing through the body impedance network, which is equivalent to omitting the second term in Equation (8.24). We retain only the current contributions associated with g_{ds} and g_{mb}. Now we get the following expression

$$I_{ds} \approx g_{ds} \cdot V_{ds} + g_{mb} \cdot V_{bs} \tag{8.26}$$

In order to eliminate V_{bs}, we apply Kirchoff's Current Law to the body node, and we obtain the following result

$$V_{bs} = V_{ds} \cdot \frac{g_{bd} + s \cdot (C_{jbd} - C_{bd})}{g_{bs} + g_{bd} + s \cdot (C_{bb} + C_{jbs} + C_{jbd})} \tag{8.27}$$

Substitute Equation (8.27) into Equation (8.26) and divide both sides by V_{ds} to obtain an expression for the output admittance.

$$y_{out} = g_{ds} + g_{mb} \cdot \frac{g_{bd} + s \cdot (C_{jbd} - C_{bd})}{g_{bs} + g_{bd} + s \cdot (C_{bb} + C_{jbs} + C_{jbd})} \tag{8.28}$$

In the low frequency (DC) limit, Equation (8.28) becomes

$$y_{out}(DC) = g_{ds} + g_{mb} \cdot \frac{g_{bd}}{g_{bs} + g_{bd}} \tag{8.29}$$

In saturation g_{bs} is at least several orders of magnitude greater than g_{bd}, since the body-drain junction is more heavily reverse biased, so we can simplify Equation (8.29) to the result obtained in [9]

$$y_{out}(DC) \approx g_{ds} \tag{8.30}$$

Note that in the below kink region, g_{bs} is typically much larger than any contribution from impact ionisation or parasitic BJT action, so we are quite justified in omitting these elements from the equivalent circuit.

In the high frequency limit, Equation (8.28) becomes

$$y_{out}(HF) = g_{ds} + g_{mb} \cdot \frac{C_{jbd} - C_{bd}}{C_{bb} + C_{jbs} + C_{jbd}} \tag{8.31}$$

In order for simulations to match the experimental frequency response shown in Figure 8.1 ($y_{out}(HF) > y_{out}(DC)$), it is necessary for the second term in Equation (8.31) to yield a positive result. Since C_{jbs}, C_{jbd} and C_{bb} are reciprocal capacitances, they all have positive signs, and g_{mb} is modelled as a positive quantity in SPICE. Therefore, the condition for qualitatively correct small-signal behaviour is given by

$$C_{jbd} > C_{bd} \tag{8.32}$$

This is an interesting result. A compact SPICE model must satisfy Equation (8.32) in order to yield the correct saturation conductance profile, yet most SPICE models for SOI do not use an integrated model for intrinsic and junction capacitances. As a result, there is no way to guarantee that C_{jbd} will always exceed C_{bd}. The incorporation of intrinsic and extrinsic capacitances into a single unified regime would be a very demanding undertaking, and will not be attempted here. Instead we will demonstrate how a straightforward empirical model for the intrinsic body charge can improve model accuracy

Empirical Body Charge Model

Body charge in the saturation region can arise in two ways; either it is the result of depletion under the front gate (in which case it accounts for part of the intrinsic body capacitance), or else it is associated with the depletion region around the drain-body pn junction (in which case it contributes to the extrinsic junction capacitance). We shall see how extrinsic capacitances are calculated in Section 8.3; for now we shall develop a basic intrinsic capacitance model, and examine how its formulation can affect the validity of a SOI compact model in the small-signal regime

Let us consider a generic scheme for assigning charge in a MOSFET device with width W and channel length L, operating in saturation. We shall employ the Gradual Channel Approximation (GCA) [12] and assume that prior to channel pinch-off, the field strength along the channel is much weaker than the vertical gate field. Closer to the drain, beyond the pinch-off point, the GCA is no longer applicable; we call this region the saturation region. The GCA region and saturation region are shown in Figure 8.3, where they are labelled as Region 1 and Region 2 respectively. Also depicted is the variation of the body charge across the two regions. Notice that in the saturation region, the intrinsic charge is expected to remain constant, since surface potential is independent of drain voltage in this region.

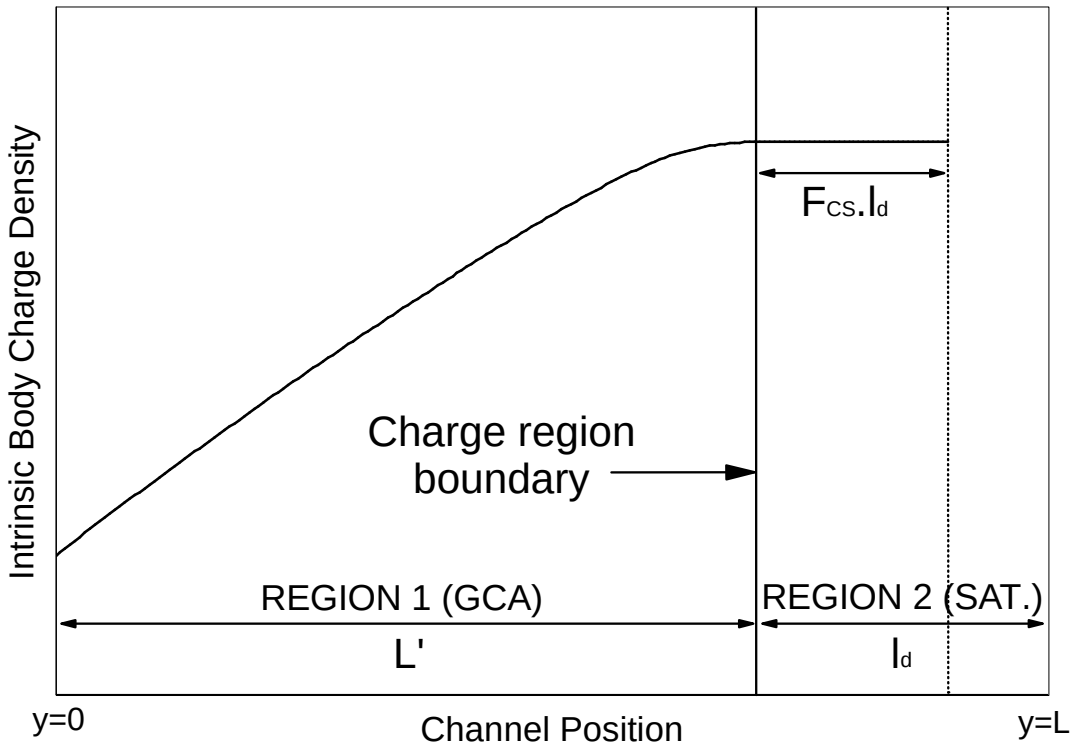

Figure 8.3: *Pictorial representation of new body charge model, showing the role of the new model parameter F_{CS}. At the boundary between Regions 1 and 2, the intrinsic body charge density is continuous.*

Q_{b1} has already been defined in Equation (8.20). Finding an expression for the drain region saturation charge Q_{B2} is more complicated, since the standard expression for q_b [12,13] is only valid in that part of the device where the Gradual Channel Approximation is valid. Because the electrical field in the saturation region has a pronounced two dimensional nature, it is difficult to derive an analytical expression which adequately describes the charge distribution in this region. In addition, some of the body charge in the saturation region will be associated with the body-drain depletion junction, and hence will not contribute to the intrinsic drain-body capacitance.

To take account of this, we introduce a new model parameter, F_{CS}, or *charge sharing factor*. F_{CS} takes a value between 0 and 1, and represents the fraction of Region 2 over which we should integrate q_b, in order to obtain Q_{B2}. The other part of the region is assumed to contain charge associated with the external junction capacitance; in effect, no intrinsic body charge exists in this region. The model is shown pictorially in Figure 8.3. In Region 2, all charge to the left of the dotted line is used to calculate Q_{B2}, and hence is associated with the intrinsic transcapacitance C_{bd}. To the right of the dotted line is the charge contributing to the external junction capacitance C_{jbd}, which is calculated using (8.41) and (8.42).

As mentioned earlier, we can assume that in the saturation region, q_b remains constant, equal to $q_b(L')$, its value at the boundary between Regions 1 and 2. According to this scheme, the expression for Q_{B2} becomes

$$Q_{B2} = W \cdot F_{CS} \cdot l_d \cdot q_b(L') \tag{8.33}$$

If we take Equation (2.18) and set $\psi_s(y)$ to ψ_{sL} (the surface potential at $y = L'$), we can use this is Equation (8.33) to obtain

$$Q_{B2} = \gamma_s C_{of} W F_{CS} l_d \left[\sqrt{\psi_{st0}} + \delta_s \left(\psi_{sL} - \psi_{st0}\right)\right] \tag{8.34}$$

Having found an expression for Q_{B2}, the next step is to obtain an expression for the intrinsic body-drain capacitance C_{bd}. Let us assume that the device is biased in saturation, and then increase the drain voltage by an infinitesimal amount dV_{ds}. According to channel length modulation theory [13], the pinch-off point for the channel will move towards the source, and so the saturation region will increase in length by an amount dL

$$dL = \lambda \cdot dV_{ds} \tag{8.35}$$

where λ is the channel length modulation (CLM) parameter given in Equation (7.20). Of course, the GCA region must be reduced in length by dL. The changes to each region are shown in Figure 8.4. The corresponding changes in the amount of charge in each region are given by

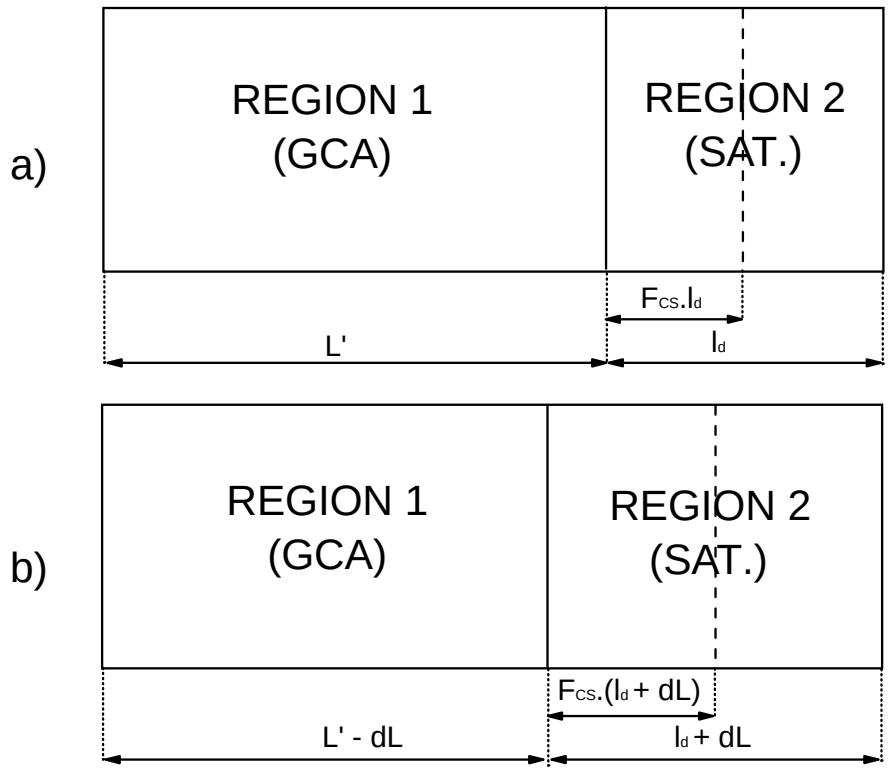

Figure 8.4: *GCA and saturation region length for a) drain voltage Vds and b) drain voltage Vds+dVds.*

$$dQ_{B1} = -W \cdot dL \cdot [q_b(L') - q_b(L' - dL)] \tag{8.36}$$

$$dQ_{B2} = W \cdot dL \cdot F_{CS} \cdot q_b(L') \tag{8.37}$$

and the total change in the body charge is obtained by summing dQ_{b1} and dQ_{b2}

$$dQ_b = W \cdot dL \cdot [(F_{CS} - 1) \cdot q_b(L') + q_b(L' - dL)] \tag{8.38}$$

From Figure 8.3 we see that the intrinsic (gate induced) depletion body charge reaches its maximum value at the boundary between Regions 1 and 2, so that

$$q_b(L') > q_b(L' - dL)] \tag{8.39}$$

Therefore, for an NMOS device (q_b is negative), dQ_{B1} will always be positive, and dQ_{B2} will always be negative. The reverse is true for a PMOS device. The sign of dQ_B is dependent on both the device and the value chosen for F_{CS}.

From the definition of intrinsic capacitance given by (8.1), we can substitute for dQ_B and dV_{ds} using (8.38) and (8.35), and obtain the following result.

$$C_{bd} = \frac{dQ_b}{dV_{ds}} = W \cdot \lambda \cdot [(F_{CS} - 1) \cdot q_b(L') + q_b(L' - dL)] \tag{8.40}$$

Evaluation of Body Charge Model

In order to evaluate the validity of the charge model, it was incorporated into STAG, and small-signal simulations of floating-body PD-SOI MOSFETs were performed. The results presented in this study are for NMOS devices. Figure 8.5 shows simulated output conductance plotted versus frequency for an NMOS device, for three values of F_{CS}. For F_{CS}=0.5 and F_{CS}=1 plausible results are obtained, with $y_{out}(HF) > y_{out}(DC)$. However, for F_{CS}=0 we find that $y_{out}(HF) < y_{out}(DC)$, so (8.32) is being violated for this case.

Let us first understand this result from a qualitative point of view. As we increase the drain voltage, the CLM effect causes the saturation region (Region 2) to grow larger, and the GCA region (Region 1) to grow smaller. Physically, we would expect some of the body charge previously in Region 1 to become associated with Region 2 instead. However, if F_{CS}=0, then the body charge in Region 2 is assumed to be zero, regardless of the size of the region, and so dQ_{b2}, which is normally a negative quantity for an NMOS device, will instead be zero. Removing body charge from Region 1, without adding any charge to Region 2, will raise dQ_b (and hence C_{bd}) to an artificially high value, making a violation of (8.32) more likely. At the other extreme, setting F_{CS}=1 will produce a much smaller rate of change of Q_b.

If we now actually substitute values of F_{CS} into (8.40), we do indeed find that the largest

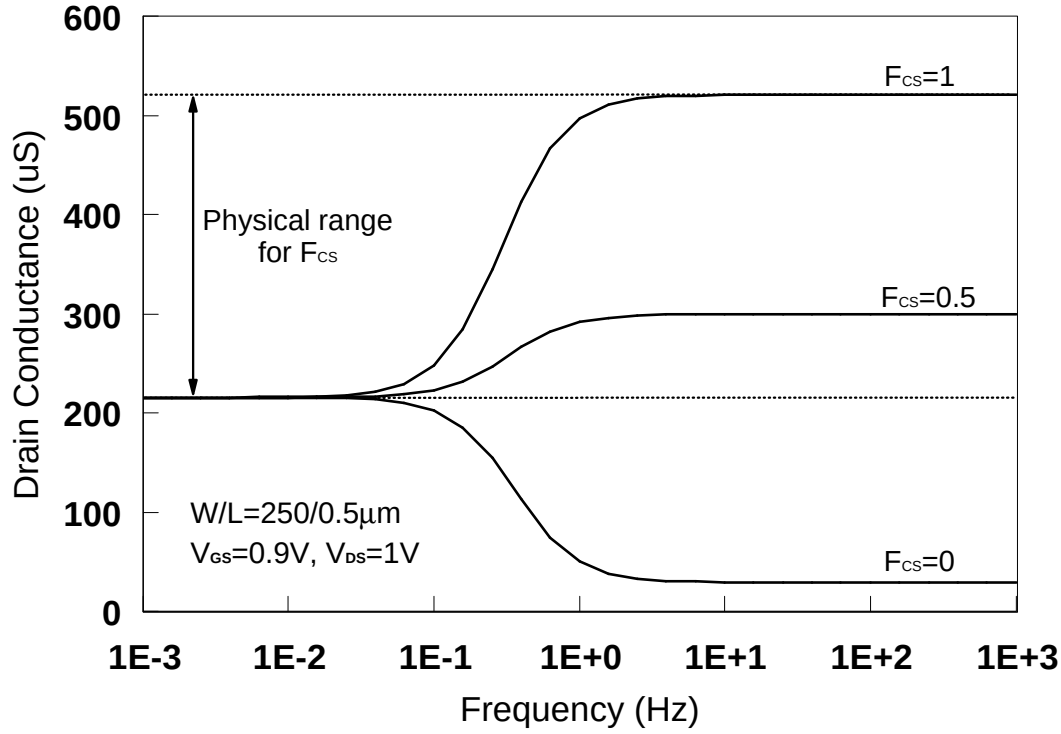

Figure 8.5: *Comparison of simulated drain conductance-frequency profiles for different values of charge sharing factor F_{CS}. The dotted lines indicate the physical range of F_{CS}, over which the correct drain conductance profile is observed.*

positive value for C_{bd} occurs when F_{CS}=0, while F_{CS}=1 actually results in a negative value for C_{bd}. In the latter case, (8.32) will always be met, but this will not necessarily yield quantitatively accurate results.

We can deduce from (8.40) that a non-physical conductance profile will be particularly evident in technologies with pronounced CLM. Conversely, simulations in which the CLM parameter, λ, is set to zero should give the correct profile, since C_{bd} will then also be zero. This is confirmed in Figure 8.6; here F_{CS} has been set to zero, and simulations have been performed using different values of λ. It can be seen that for a combination of a high λ value and a low value for F_{CS}, the simulated conductance becomes negative, even though self-heating has been switched off for these simulations. This occurs when the negative contribution from the second term in (8.31) becomes larger than g_{ds}, and is a clear indication that the model is yielding non-physical results.

F_{CS}=0 and F_{CS}=1 correspond to the lower and upper limits on the possible range of high frequency drain conductance values. However, since we still get non-physical behaviour for low values of F_{CS}, the physical lower limit for F_{CS} actually corresponds to the value required to make the $y_{out}(HF)$ equal to g_{ds}. We would expect the optimum value for F_{CS} to lie somewhere between this lower limit and F_{CS}=1; the physical range is indicated

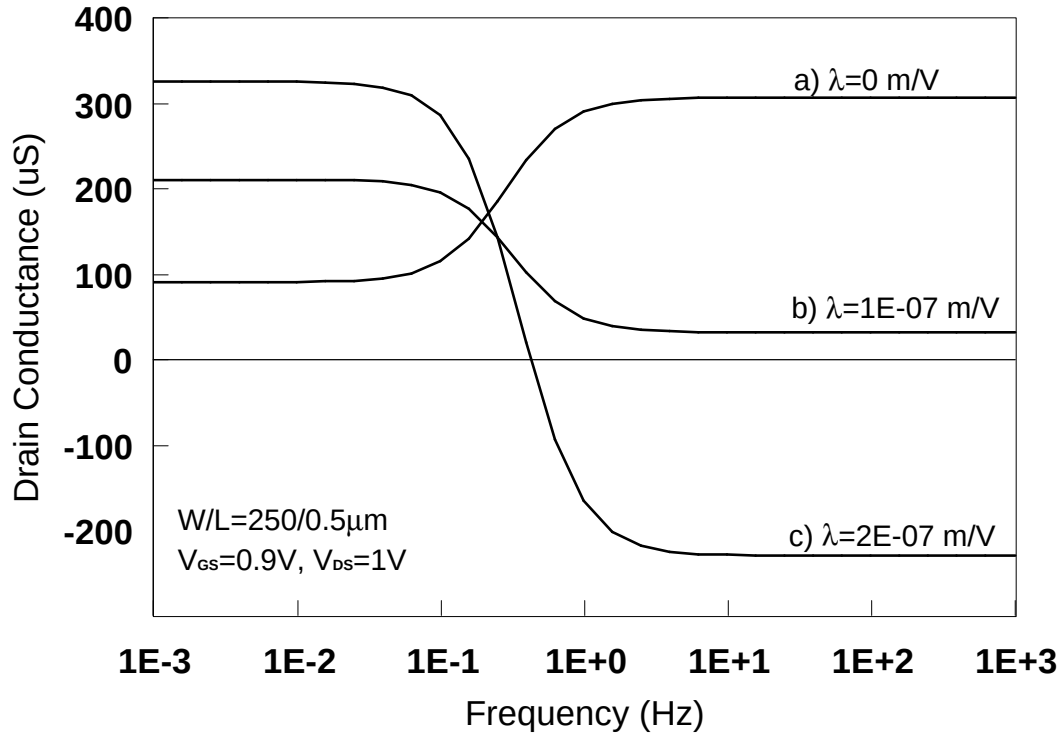

Figure 8.6: *Variation of simulated drain conductance with channel length modulation parameter λ, for F_{CS}=0. When λ is set to zero, a physically correct drain conductance profile is observed.*

on Figure 8.5. When applying this model to all device instances of a particular process technology, an averaged value for F_{CS} will probably be required, bearing in mind that the degree of charge sharing is likely to vary for different device dimensions and bias conditions.

We can attempt to fit simulation results to data obtained from measurements, to see if experimental values of drain conductance fall within the acceptable range defined in Figure 8.5. Such a fit is shown in Figure 8.7, using a device manufactured using Technology B (see Appendix A). A value of 0.5 was chosen for F_{CS}. The simulated data (solid lines) were obtained using the same parameter set used in Figure 8.5, from which we can see that F_{CS}=0.5 does indeed fall within the allowed range of values. Using this value, we can see that reasonable matching is obtained, and that the below-kink saturation conductance remains close to the measured curve in the region of interest.

If we now focus specifically on the below-kink saturation region, we can better gauge the usefulness of the new model. Figure 8.8 shows the same plot as Figure 8.7, but with the scale adjusted. Three sets of simulation data are displayed, corresponding to F_{CS} values of 0, 0.5, and 1. Values of 0 and 1 yield conductances which are too low and too high respectively, whereas F_{CS}=0.5 does give a reasonable result over the relevant voltage range. Because of minimal fitting of the DC model for these simulations, the onset into satura-

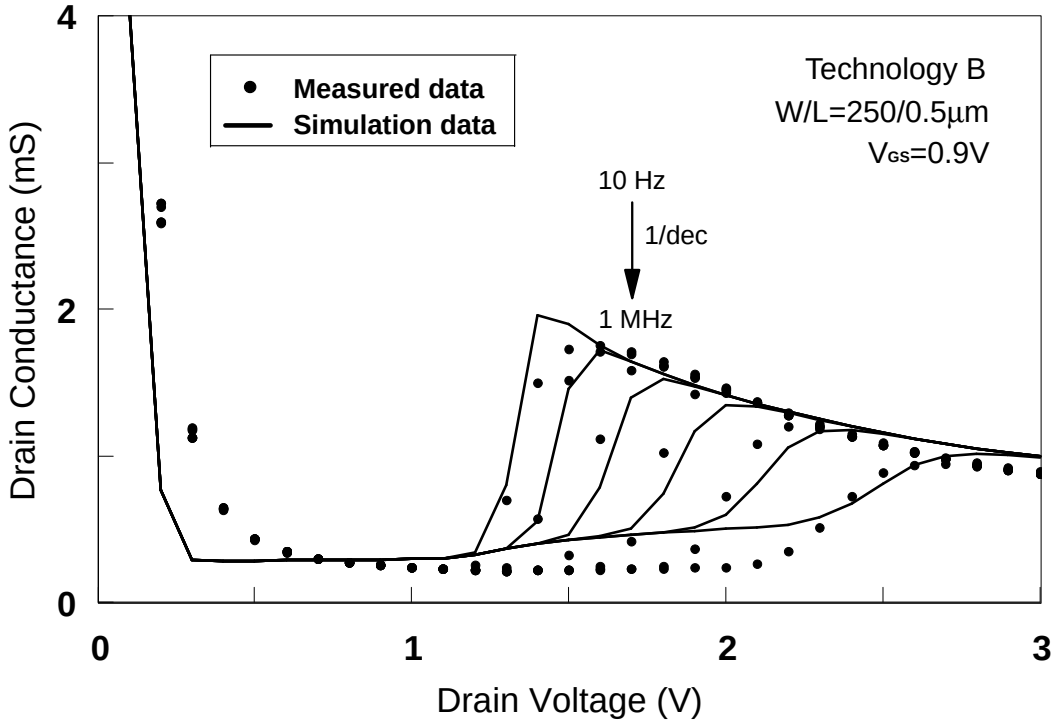

Figure 8.7: *Simulated small-signal drain conductance versus drain bias, compared with measured data. New charge sharing factor is used to model saturation body charge, with F_{CS} set to 0.5 for the simulation data.*

tion is rather abrupt. Additionally, the fact that F_{CS} is modelled as a constant, without any voltage dependence, means that fitting is not perfect in the saturation region either. Clearly the new body charge model would benefit from further development, but it is at least sufficiently good to demonstrate the physical principles underlying the small-signal behaviour.

The evaluation results presented in this section show that it is possible to greatly improve the accuracy of the small-signal modelling of SOI devices, using a simple theory, and requiring only one additional parameter in addition to the basic DC parameter set. The basic concept of using a charge sharing factor appears sound, and although more complicated schemes could be devised in order to better reflect the terminal voltage dependency of F_{CS}, the empirical approach described in this work has been shown to yield adequate small-signal characteristics. A more robust solution would be the implementation of an integrated capacitance model, in which extrinsic and intrinsic capacitances are interdependent. This would help ensure that the relative magnitudes of the two contributions satisfied the condition for qualitatively correct conductance profiles, and ideally would be able to provide quantative conductance matching as well. However, the author is not aware of any such capacitance model suitable for use in a compact model.

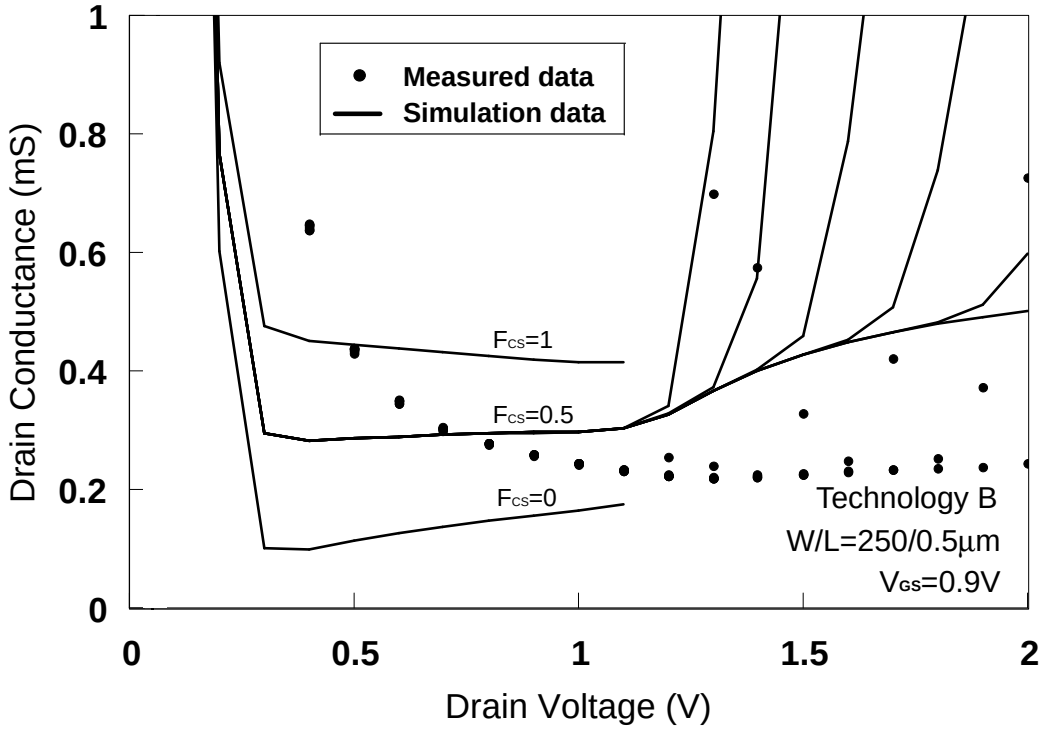

Figure 8.8: *Plot of the small-signal drain conductance versus drain bias, focusing on the below-kink saturation region. Measured data is compared with simulation data for F_{CS}=0, 0.5, and 1.*

Influence of Minimum Nodal Conductance on SOI AC Simulations

We shall conclude this section by discussing a related issue, which if not handled carefully, can also result in non-physical characteristics. In many SPICE-based compact MOSFET models, an additional conductance g_{MIN} is placed in parallel with the junction conductances. SPICE uses g_{MIN} to provide a minimum nodal conductance, in order to ensure that there is a non-zero conductance between these pairs of nodes in the equivalent circuit. SPICE can step up the value of g_{MIN} in order to help individual devices to converge. It is not a physical conductance, so when convergence is not an issue, one must be wary of assigning to it a value which is comparable with any physical conductance.

For most bulk MOSFET modelling applications, a default value for g_{MIN} of 10^{-12}S is found to be acceptable. However, in the case of SOI, using a default value of that magnitude can drastically alter the simulated conductances above and below the low frequency pole. This is because most of the body network impedances in a SOI device are usually smaller than 10^{-12}S. To illustrate this, Figure 8.9 shows the results of small-signal simulations using two different values of g_{MIN}. Below 10^{-25}S, there is found to be no significant change in conductance, suggesting that this value is small enough to be used in SOI simulations. When we set g_{MIN} to 10^{-12}S, we can see that the difference in drain conductance is very noticeable. This occurs when g_{MIN} becomes comparable to g_{bs}, so

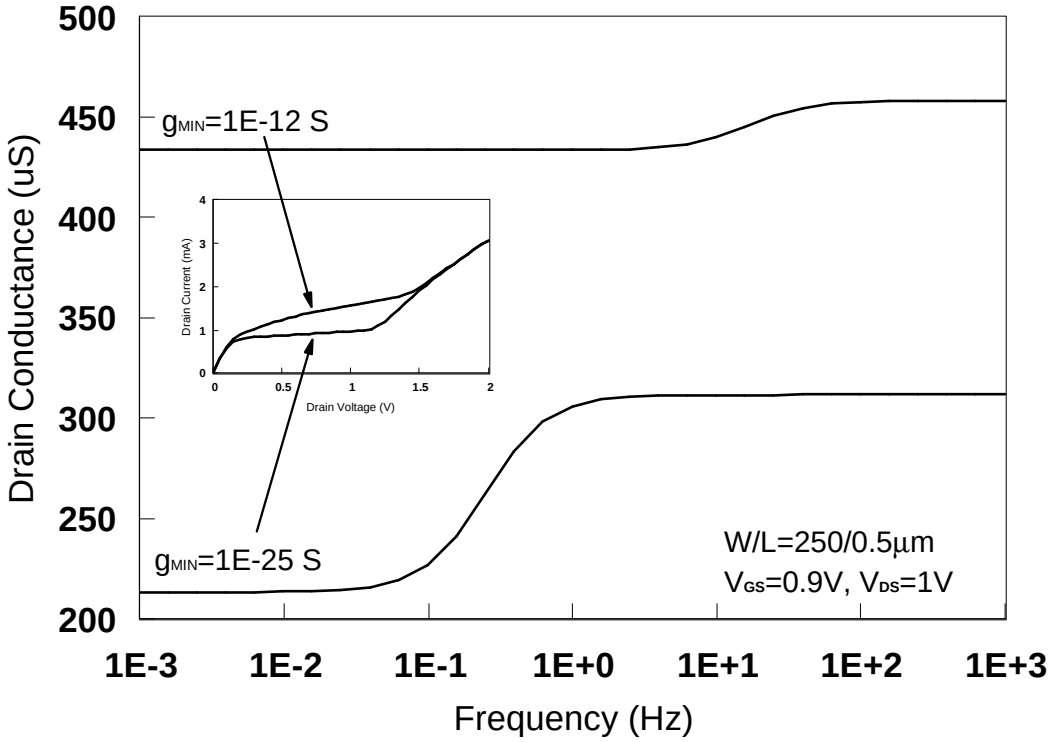

Figure 8.9: *Simulated drain conductance versus frequency, for $g_{MIN} = 10^{-25}$ and $10^{-12}S$. The inset shows simulated drain current characteristics for the same device and conditions, showing anomalous effect on saturation characteristics of making g_{MIN} comparable in magnitude to physical body conductances.*

that a significant proportion of the drain-source voltage is dropped between the body and source nodes. When this happens, the threshold voltage is reduced by means of the body effect, thus altering the DC operating point. From the inset of Figure 8.9, we can see how g_{MIN} can affect DC simulation results as well as AC.

We conclude that a much smaller value of g_{MIN} must be employed in SOI simulations. A value of 10^{-25}S has been found to be suitable, since this will make g_{MIN} at least one or two orders of magnitude smaller than the drain-body junction conductance g_{bd} under most conditions. Setting g_{MIN} to such a small value should be done *as a matter of routine*, to ensure accurate floating body SOI simulations, and to avoid confusion when trying to distinguish this kind of simulator anomaly from problems inherent to the compact model itself.

8.3 Extrinsic Capacitances

Extrinsic capacitances are usually modelled using separate auxilliary equations, which exist independently of the main charge model. They are usually empirical expressions intended to provide a first-order model, one which yields reasonably accurate values over most operating conditions, without recourse to complicated mathematical schemes. Of greatest relevance to this work are the body-source and body-drain junction capacitances, C_{jbs} and C_{jbd}, which can be modelled using pn junction theory. A widely used scheme [12, 14] is to calculate C_{jbs} and C_{jbd} from C_{J0}, the capacitance per unit sidewall area

$$C_{jbd} = W \cdot t_b \cdot \frac{C_{J0}}{\left(1 - \frac{Vbd}{PB}\right)^M} \quad if\ V_{bd} < FC \cdot PB \tag{8.41}$$

$$C_{jbd} = W \cdot t_b \cdot \frac{C_{J0}}{(1 - FC)^{(1+M)}} \left[1 - FC\,(1 + M) + \frac{V_{bd}}{PB} M\right] \quad if\ V_{bd} \geq FC \cdot PB \tag{8.42}$$

where FC, PB, and M are empirical parameters defined in [14]. A similar expression can be obtained for the source-body junction capacitance, by replacing V_{bd} with V_{bs}.

8.3.1 Source and Drain Junction Depletion Capacitance

The model priorities for calculating drain-body and source-body junction depletion capacitances, C_{jbd} and C_{jbs} are given below:

1. By directly specifying values for C_{jbd} and C_{jbs} in the model netlist, the user overrides all other priorities.

2. The user can provide a value for C_J, the junction capacitance per unit area of junction sidewall. STAG will then calculate C_{jbd} and C_{jbs} from C_{J0}.

3. If neither C_{jbd}/C_{jbs} or C_J are provided, then STAG3 will calculate a default value for C_J using the following relation:

$$C_J = \sqrt{\frac{\epsilon_{si} q}{2\phi_{bi}} \left(\frac{N_B \cdot N_{HDD}}{N_B + N_{HDD}}\right)} \tag{8.43}$$

where N_B is the silicon film doping concentration, N_{HDD} is the source/drain region doping concentration (defaults to $10^{20} cm^{-3}$), and ϕ_{bi} is the thermal equilibrium voltage across the pn junction, defined as

$$\phi_{bi} = \phi_t \ln\left(\frac{N_B \cdot N_{HDD}}{n_i^2}\right) \tag{8.44}$$

If the need arises to switch off the junction depletion capacitances (for instance to test the intrinsic capacitance model without interference from other capacitance contributions) then C_J can be set to zero in the model netlist.

8.3.2 Overlap Capacitances

In STAG3, all overlap capacitances are assigned default values, although these can be overridden by setting parameter values in the netlist. The following set of equations are used to calculate the default values.

$$C_{gfso} = \frac{\epsilon_0 \cdot \epsilon_{ox} \cdot l_{diff} \cdot W}{t_{of}} \tag{8.45}$$

$$C_{gfdo} = \frac{\epsilon_0 \cdot \epsilon_{ox} \cdot l_{diff} \cdot W}{t_{of}} \tag{8.46}$$

$$C_{gfbo} = \frac{\epsilon_0 \cdot \epsilon_{ox} \cdot (L - 2 \cdot l_{diff}) \cdot (0.1 \cdot 10^{-6} \cdot F_{min})}{t_{of}} \tag{8.47}$$

$$C_{gbso} = \frac{\epsilon_0 \cdot \epsilon_{ox} \cdot (2 \cdot 10^{-6} \cdot F_{min} + l_{diff}) \cdot W}{t_{ob}} \tag{8.48}$$

$$C_{gbdo} = \frac{\epsilon_0 \cdot \epsilon_{ox} \cdot (2 \cdot 10^{-6} \cdot F_{min} + l_{diff}) \cdot W}{t_{ob}} \tag{8.49}$$

$$C_{gbbo} = \frac{\epsilon_0 \cdot \epsilon_{ox} \cdot (L - 2 \cdot l_{diff}) \cdot (0.1 \cdot 10^{-6} \cdot F_{min} + W)}{t_{ob}} \tag{8.50}$$

where l_{diff} is a model parameter specifying the lateral diffusion distance of the source and drain regions under the front gate. Note that the units of F_{min} are microns (μm), so a factor of 10^{-6} is needed to convert to metres.

Front gate overlap capacitances can be set instead by specifying CGFSO, CGFDO, and/or CGFBO. This can be used to switch off certain capacitances, by setting the appropriate model parameter to zero. Back gate overlap capacitances are calculated per unit area. Therefore, to set back gate overlap capacitances in the netlist, it is necessary to specify CGBSO/CGBDO/CGBBO, and also $A_S/A_D/A_B$, the contact areas for the source, drain and body regions. The exception to this is if it is necessary to set one or more of the back gate overlap capacitances to zero. In this case, all that is required is to set CGBSO, CGBDO, and/or CGBBO to zero in the netlist.

In the event that netlist values are used, the overlap capacitances are calculated according to the following set of equations

$$C_{gfso} = CGFSO \cdot W \tag{8.51}$$

$$C_{gfdo} = CGFDO \cdot W \tag{8.52}$$

$$C_{gfbo} = CGFBO \cdot (L - 2 \cdot l_{diff}) \tag{8.53}$$

$$C_{gbso} = CGBSO \cdot A_S \tag{8.54}$$

$$C_{gbdo} = CGBDO \cdot A_D \tag{8.55}$$

$$C_{gbbo} = CGBBO \cdot A_B \tag{8.56}$$

Note that any combination of overlap capacitances can be specified or be left as default. For instance, setting CGFSO, CGBBO and A_B would set front gate-source and back gate-body overlap capacitances, with the other four using their default values.

References

[1] K A Sakallah, Y T Yen, and S S Greenberg, "A first-order charge conserving MOS capacitance model", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 9, no. 1, pp. 99–108, Apr. 1990.

[2] J.E. Meyer, "MOS Models and Circuit Simulation", *RCA Review*, vol. 32, pp. 42–62, 1971.

[3] R Howes, W Redman-White, K G Nichols, P J Mole, M J Robinson, and S Bird, "An SOS MOSFET model based on calculation of the surface potential", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 6, pp. 494–506, Apr. 1994.

[4] D E Ward and R W Dutton, "A Charge-Oriented Model for MOS Transistor Capacitances", *IEEE Journal of Solid-State Circuits*, vol. SC-13, pp. 703–708, 1978.

[5] Colin C. McAndrew, and JamesJ. Victory, "Accuracy of Approximations in MOSFET Charge Models", *IEEE Transactions on Electron Devices*, vol. 49, no. 1, pp. 72–81, 2002.

[6] C F Edwards, B M Tenbroek, M S L Lee, W Redman-White, and M J Uren, "The effect of body contact series resistance on SOI CMOS amplifier stages", *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2290–2294, Dec. 1997.

[7] J Tihanyi and H. Schlotterer, "Properties of EFSI MOS transistors due to the floating substrate and the finite volume", *IEEE Transactions on Electron Devices*, vol. ED-22, pp. 1017–1023, 1975.

[8] S S Eaton and B Lavevic, "The effect of a floating substrate on the operation of silicon-on-sapphire transistors", *IEEE Transactions on Electron Devices*, vol. ED-25, pp. 907–912, Aug. 1978.

[9] R Howes and W Redman-White, "A small-signal model for the frequency-dependent drain admittance in floating-substrate MOSFETs", *IEEE Journal of Solid-State Circuits*, vol. 27, no. 8, pp. 1186–1193, Aug. 1992.

[10] W Redman-White, B Tenbroek, M S L Lee, C F Edwards, M J Uren, and R J T Bunyan, "Analogue design issues for SOI CMOS", in *Proceedings IEEE international SOI conference*, 1996, pp. 6–8.

[11] H C de Graaff and F M Klaassen, *Compact Transistor Modelling for Circuit Design*, Springer-Verlag, 1990.

[12] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[13] Y P Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill Book Company, 1987.

[14] D A Divekar, *FET Modeling for Circuit Simulation*, Kluwer Academic Publishers, 1988.

Chapter 9

Model Evaluation Results

9.1 Introduction

In order to test the fitting capabilities of STAG3, single device characteristics are simulated and compared with those measured from a commercial $0.25\mu m$ PD-SOI process technology. Due to time limitations, only DC results for body-tied and floating NMOS devices are given here. It would of course have been preferable to include other results: results for PMOS characteristics, charge plots, and miscellaneous others. It should be pointed out however that most of the important changes to STAG3 have not involved changes to the floating body or charge models; these models were developed for STAG2, and provided good fitting for older technologies down to about $0.7\mu m$ [1]. The key to achieving good fitting for STAG for deep-submicron devices have been the DC core and high field model improvements, and to evaluate these we can rely on DC characteristics.

Unless otherwise stated, all results in this chapter are for Technology A (see Appendix A). The default self-heating models have been used, so that the active thermal area is calculated from the device dimensions, and used to calculate R_T and C_T. The associated model parameter set used to simulate Technology A is given in Table 9.1 at the end of this chapter.

9.2 Transconductance Results

To provide an initial idea of STAG3's accuracy, and to extract and adjust some basic model parameters, drain current was plotted as a function of gate voltage, for both low and high drain voltages. As with most of the evaulations in this chapter, results have been obtained for both a long channel device ($W = 10\mu m$, $L = 10\mu m$) and a short channel device ($W = 10\mu m$, $L = 0.25\mu m$), with the same model parameter set being used to generate all results. Fig. 9.1 shows the long channel data.

It can be seen that the match to experimental data is quite good, with only a small number of parameters needing extraction and optimisation in order to obtain these results. The extracted threshold voltage was used, in conjunction with the model parameters from

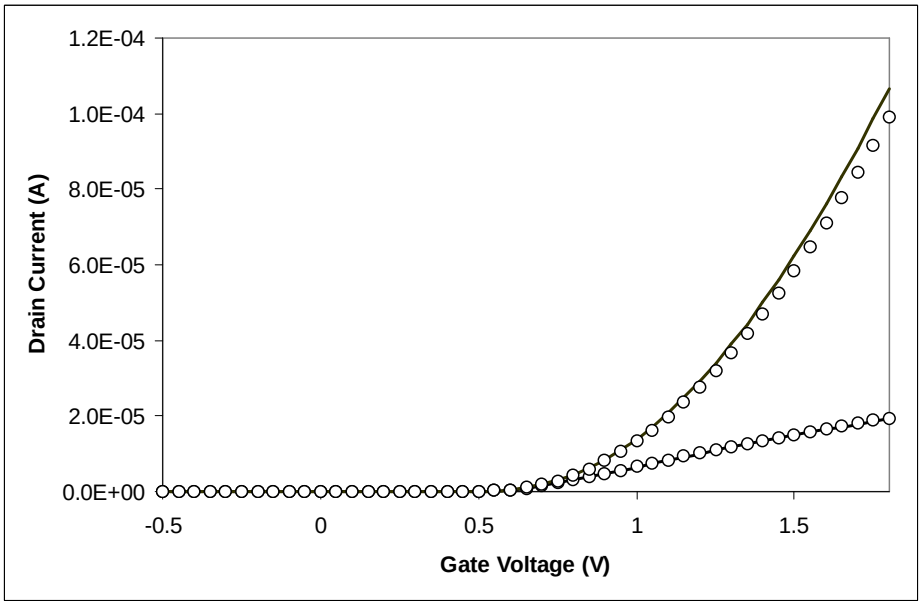

Figure 9.1: *Drain current versus gate voltage for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 10\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

the threshold fitting model in Section 6.7.3. The mobility was extracted, and the surface roughness parameter was fitted to the portion of the curve corresponding to high V_{GfS}, since surface roughness is the dominant scattering process in this region. The Coulomb and phonon scattering parameters were then hand fitted. In order to get better fitting for the high drain voltage curve, the DIBL parameter σ was adjusted by hand. The front gate thickness and a uniform doping concentration were obtained from process data.

Next we examine the transconductance curves for the same device (Fig. 9.2). Along with the output conductance, the transconductance determines the gain of simple amplifier stages, making it an important parameter for analogue circuit designers.

While good accuracy is seen for $V_{DS} = 0.1V$, there is a noticeable deviation for $V_{DS} = 1.8V$. This is probably due to inaccuracies caused by the linearisation of the channel and body charges around the source. As we saw in Chapter 2, this can result in problems correctly estimating the channel charge density, and hence the drain current, at high drain biases.

Although this was not a priority during this work, we shall also look at the higher derivatives of drain current with respect to the gate voltage, since these come into play during distortion analysis [2]. In Figs. 9.3 and 9.4, we plot the 2nd and 3rd order derivatives of the drain current. It can be seen that STAG3 reproduces all of the qualitative features, and

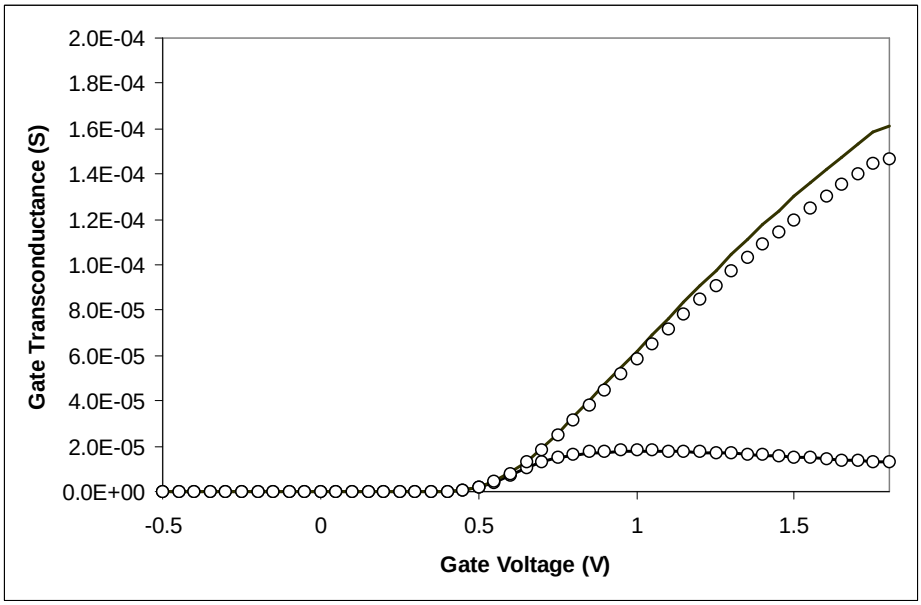

Figure 9.2: *Gate transconductance versus gate voltage for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 10\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

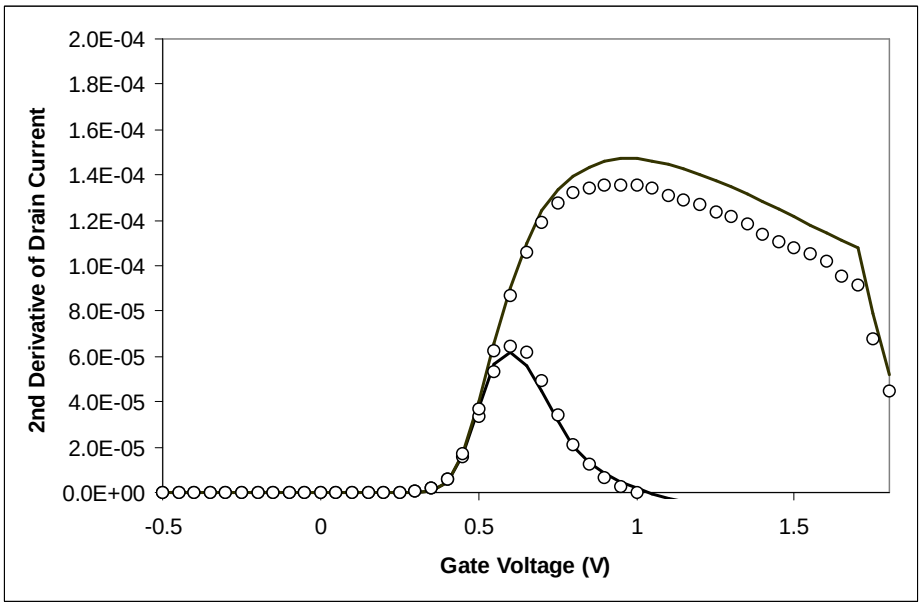

Figure 9.3: *2nd derivative of drain current w.r.t. gate voltage versus gate voltage for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 10\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

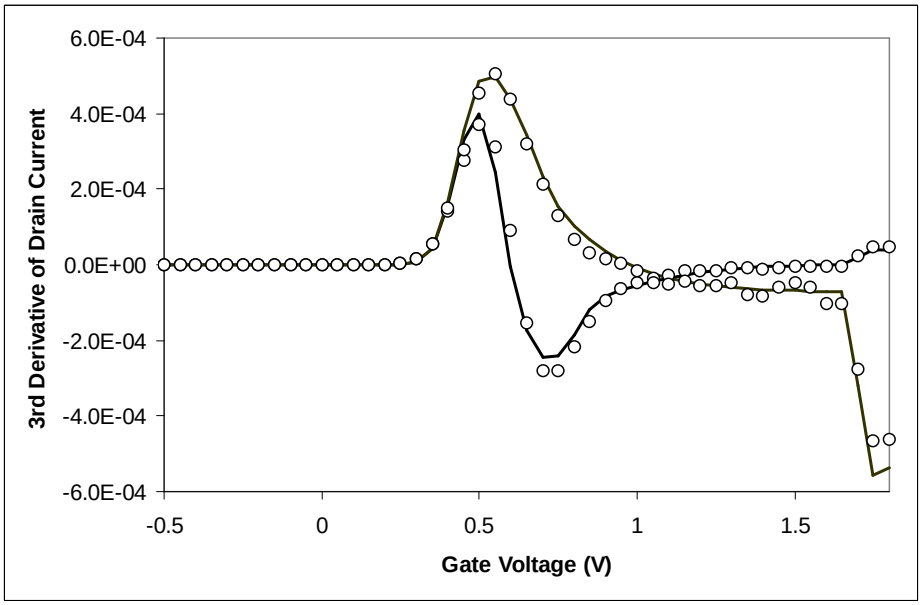

Figure 9.4: *3rd derivative of drain current w.r.t. gate voltage versus gate voltage for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 10\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

provides high accuracy for quantitative matching as well. Some deviations are again seen for $V_{DS} = 1.8V$, but again we can put these down to problems due to charge linearisation.

Looking now at the short channel device, we see that a similar degree of accuracy has been obtained for the drain current as it was for the longer device (Fig. 9.5). The same parameters were used, and in addition the short channel effect parameter was extracted. The RSCE parameters were then hand fitted to give a good estimation of the threshold voltage.

Getting close fits for the measured transconductance data was much more problematic, as can be seen in Fig. 9.6. Once again, the problem is worse at high V_{DS}. Part of the problem can once again be attributed to issues arising from the STAG3 charge linearisation. However, other parts of the model might be contributing to this error.

For instance, as was discussed in Section 3.3.4, there are some aspects of the mobility model which could be improved to be more physical. The use of Matthieson's rule is one such aspect, though it is unclear exactly how much accuracy is being lost through its use. Furthermore, no obvious alternative is available; hence its popularity in compact models. One omission which can be assumed to have an impact is the absence of inter-valley scattering. As was mentioned in Section 3.3.4, this is generally neglected in compact models, as it is a complicated function of the gate voltage. Whether there is also a significant dependence on drain voltage is unclear, but it seems plausible. Either way, these kinds of

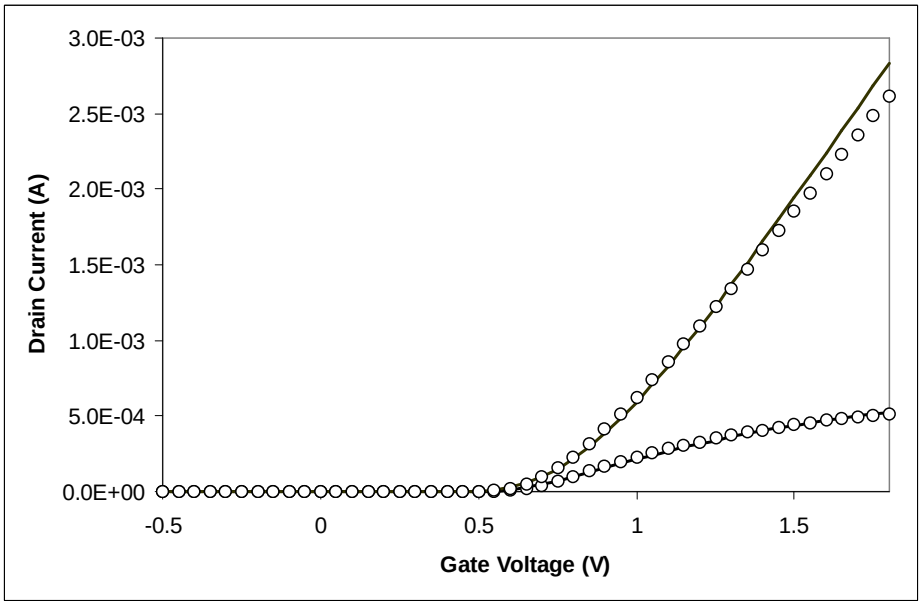

Figure 9.5: *Drain current versus gate voltage for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 0.25\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

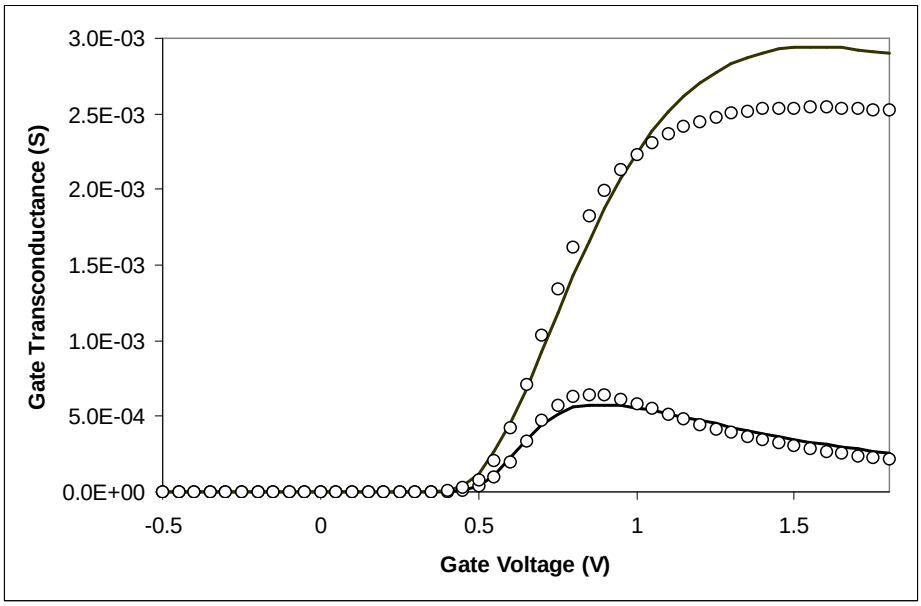

Figure 9.6: *Gate transconductance versus gate voltage for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 0.25\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

simplications can be expected to reduce the overall ability of the model to fit to transconductance data over a range of biases and channel lengths.

Another possible source of error might come from the way that the drain series resistance is being modelled. No significant relation was found between series resistance and gate voltage for this process technology, due to the lack of LDD regions. However, STAG3 lacks the ability to model the resistance associated with the saturated drain region; this has to be included as part of either the internal or external series resistance models. This could be modelled in STAG3 by increasing the drain series resistance to be larger than the source value. However, this measure alone is insufficient to give a good fit to the data in Fig. 9.6, and so the two series resistances have been left equal in these simulations.

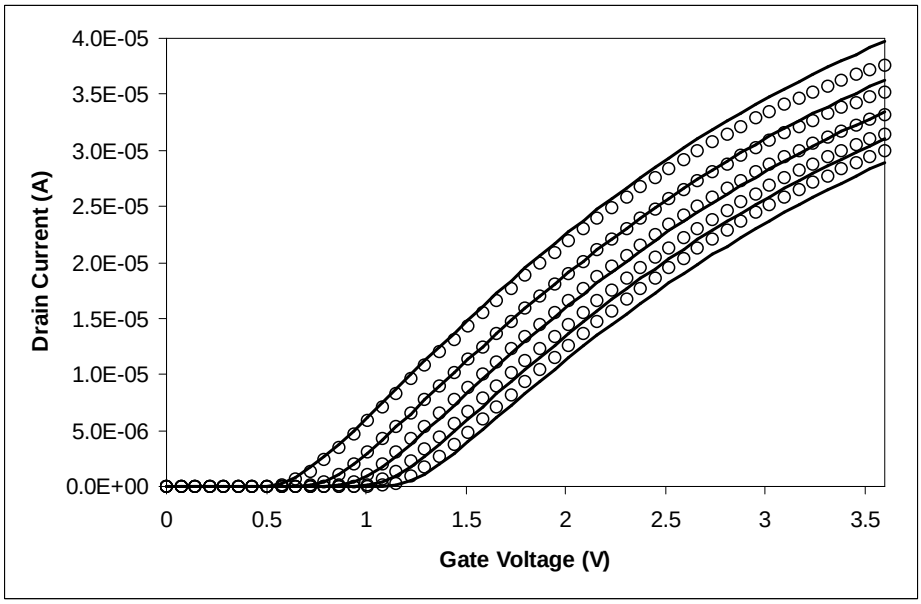

Figure 9.7: *Drain current versus gate voltage for $V_{DS} = 0.1V$ and $V_{BS} = 0$, -0.5, -1.0, -1.5, -2.0, and $-2.5V$ for a $W = 50\mu m$, $L = 50\mu m$ body tied device in Technology B. Circles indicate measured values, solid lines indicate simulated results.*

Before we conclude this section, we should examine the ability of STAG3 to model variations in the body-source voltage V_{BS}. Because we lack the appropriate experimental data for Technology A, we shall perform our comparisons using measured data from Technology B (see Appendix A). Figs. 9.7 and 9.8 show how transconductance characteristics change as V_{BS} is made increasingly negative and the threshold voltage increases. It can be seen that the matching across the whole V_{BS} range is good, which indicates that the STAG3 body model is sufficiently accurate.

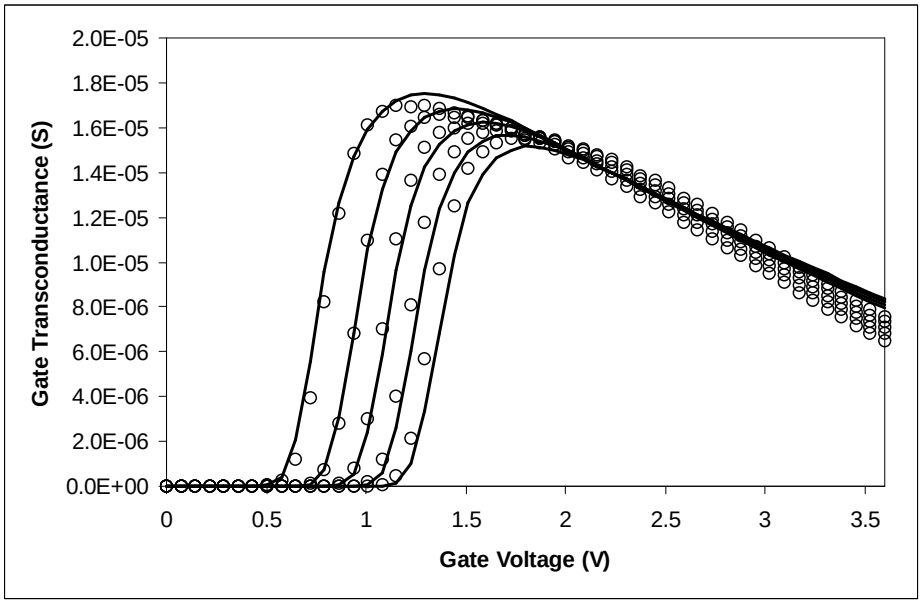

Figure 9.8: *Gate transconductance versus gate voltage for $V_{DS} = 0.1V$ and $V_{BS} = 0, -0.5, -1.0, -1.5, -2.0,$ and $-2.5V$ for a $W = 50\mu m$, $L = 50\mu m$ body tied device in Technology B. Circles indicate measured values, solid lines indicate simulated results.*

9.3 Subthreshold Results

The subthreshold region is important because it determines the leakage current and therefore the power dissipation of the device in the OFF state. This is of particular interest for modern low-power applications. The measured and simulated subthreshold characteristics are compared in Figs. 9.9 and 9.10.

The matching is good in most regions, and while there are some inaccuracies in the subthreshold slope, this is probably a result of not knowing the body doping profile for this technology. If it were known, then it would be possible to use the STAG3 non-uniform doping model to improve the fitting in this region.

The poorest matching comes in the region of negative gate voltage. Not only is the STAG surface potential model not designed to give accurate results in this region, but it also lacks the ability to model Gate Induced Drain Leakage (GIDL). GIDL can be seen in both sets of measured data, but particularly in the short channel results. It manifests as an increase in subthreshold current as the gate voltage continues to be reduced, and is caused by strong drain-gate electric fields developing near the drain region. More will be said about this in Chapter 10.

The subthreshold leakage model parameters were optimised to give an acceptable fit in

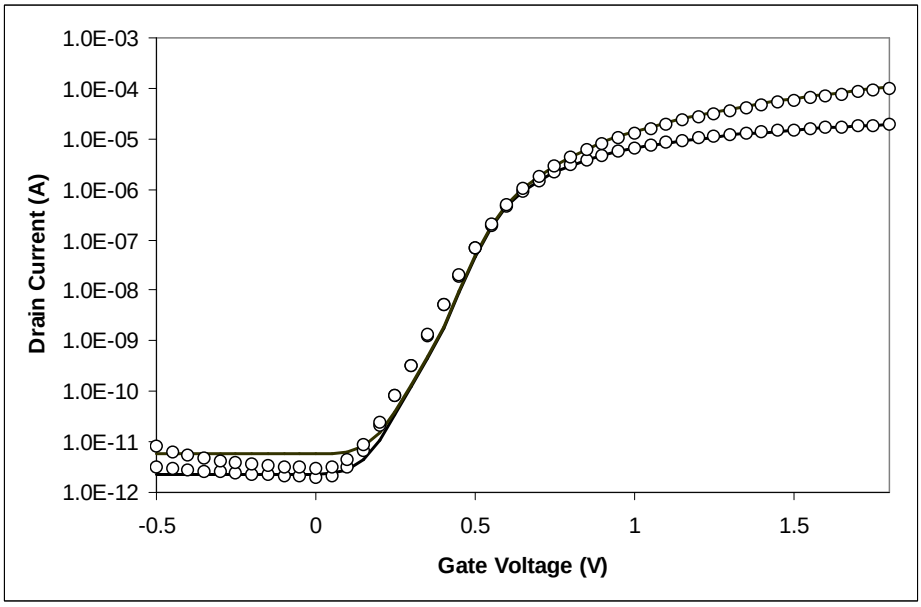

Figure 9.9: *Subthreshold characteristics for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 10\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

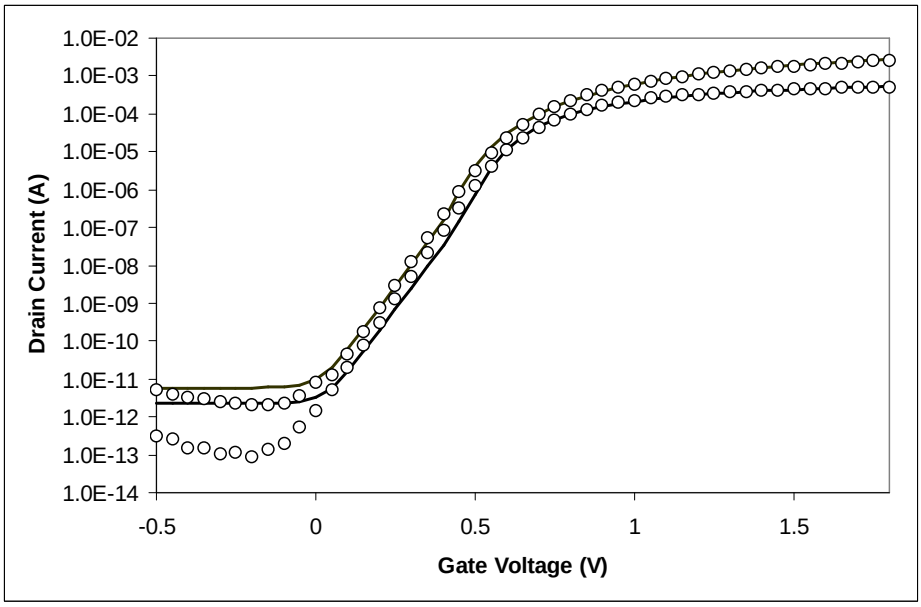

Figure 9.10: *Subthreshold characteristics for $V_{DS} = 0.1V$ and $1.8V$ for a $W = 10\mu m$, $L = 0.25\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

the long channel case, and no attempt was made to improve the fit for the short channel data.

9.4 Output Characteristics

For the drain voltage characteristics, an additional number of parameters were optimised. These included the external source and drain series resistances and the saturation velocity v_{sat}; none of these parameters deviated much from their extracted or default values). Aditionally, having used these parameters to fit below the onset of saturation, the sub-micron CLM model was hand fitted to improve accuracy in the saturated region.

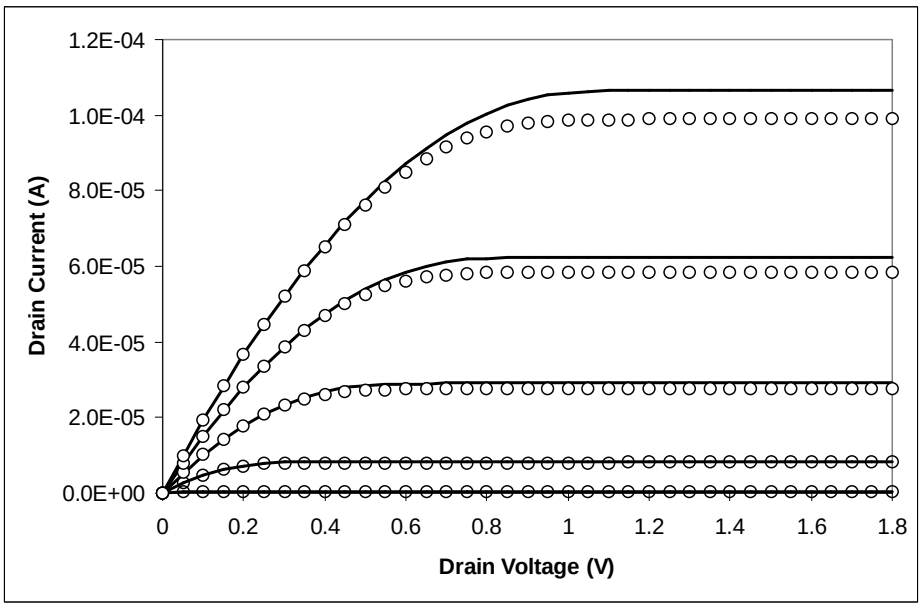

Figure 9.11: *Drain current versus drain voltage for $V_{GfS} = 0$, 0.3, 0.6, 0.9, 1.2, 1.5, and $1.8V$ for a $W = 10\mu m$, $L = 10\mu m$ body tied device. Circles indicate measured values, solid lines indicate simulated results.*

The resulting simulations can be seen in Figs. 9.11 and 9.12. Overall the fitting is good, both in terms of current levels in saturation, and the onset of saturation itself. Some overestimation of the saturation current occurs for both long and short channel devices, at high gate voltages. In many analogue circuits, devices are biased with gate overdrives of $0.5V$ or less, and so the model parameters have been hand fitted to ensure better fitting in this region, at the expense of the high gate voltage range.

Figs. 9.13 and 9.14 show the corresponding plots of output conductance against drain voltage. Output conductance is another important parameter for analogue designers. In STAG2, there was a problem that the conductance plots entered into saturation too

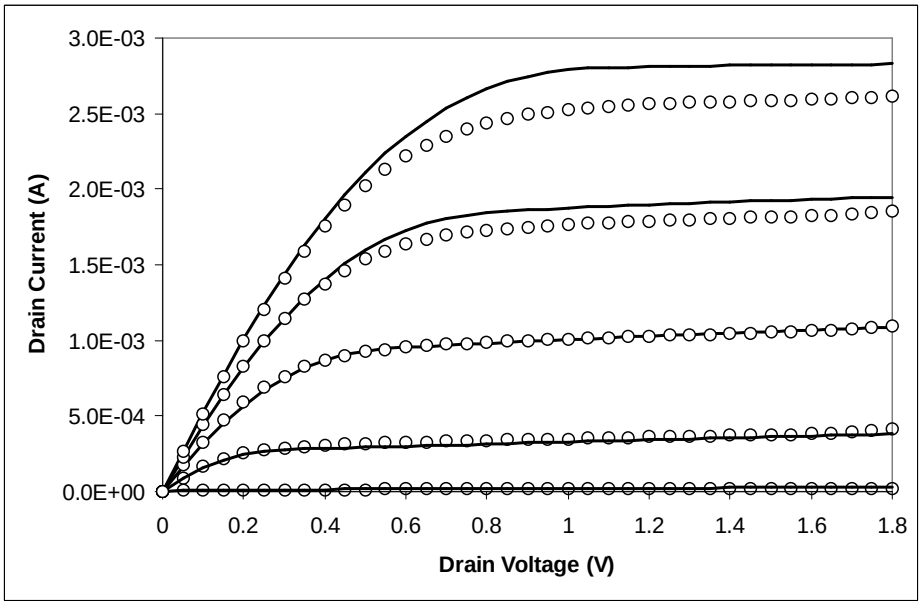

Figure 9.12: *Drain current versus drain voltage for* $V_{GfS} = 0$, 0.3, 0.6, 0.9, 1.2, 1.5, *and* $1.8V$ *for a* $W = 10\mu m$, $L = 0.25\mu m$ *body tied device. Circles indicate measured values, solid lines indicate simulated results.*

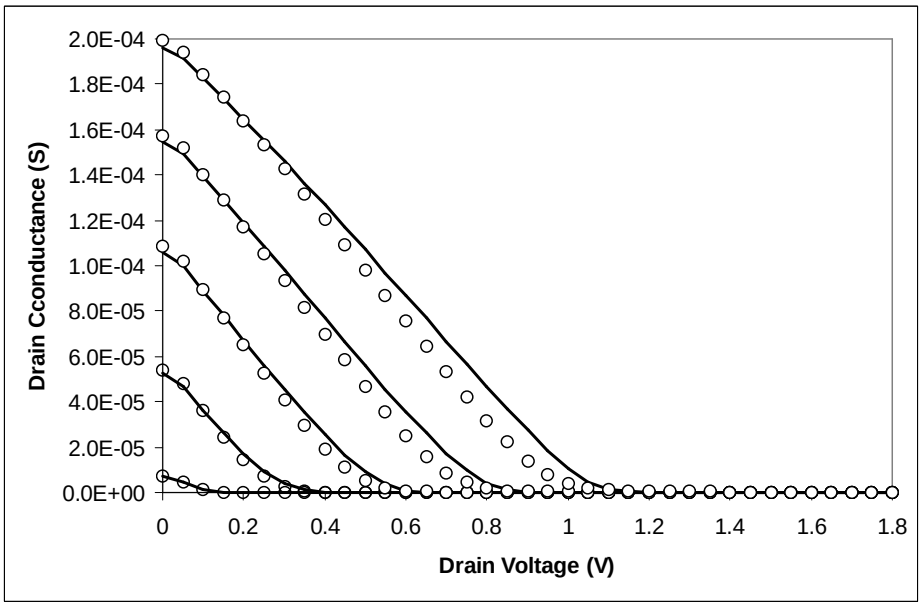

Figure 9.13: *Output conductance versus drain voltage for* $V_{GfS} = 0$, 0.3, 0.6, 0.9, 1.2, 1.5, *and* $1.8V$ *for a* $W = 10\mu m$, $L = 10\mu m$ *body tied device. Circles indicate measured values, solid lines indicate simulated results.*

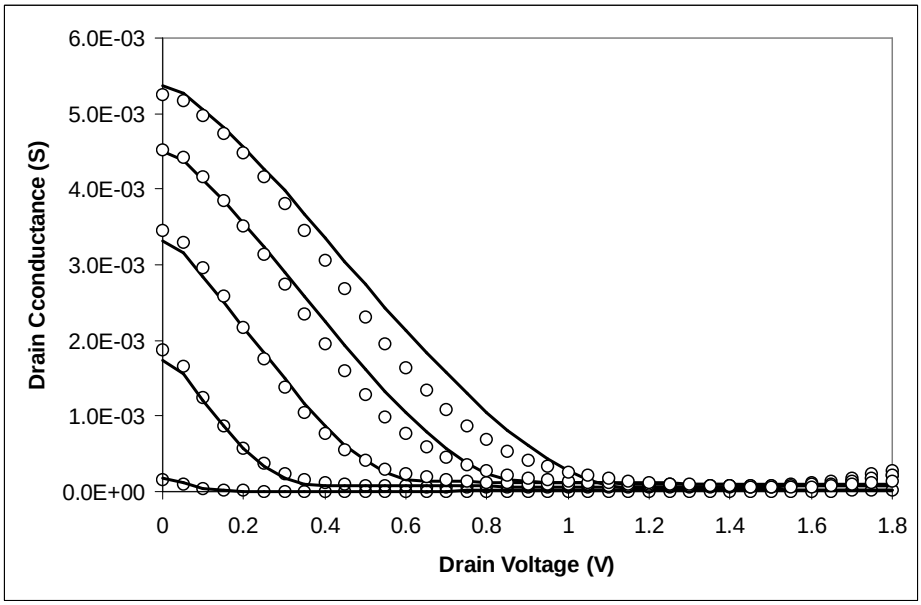

Figure 9.14: *Output conductance versus drain voltage for* $V_{GfS} = 0$, 0.3, 0.6, 0.9, 1.2, 1.5, *and* $1.8V$ *for a* $W = 10\mu m$, $L = 0.25\mu m$ *body tied device. Circles indicate measured values, solid lines indicate simulated results.*

abruptly [1], a serious issue since designers need to know the onset quite exactly in order to bias devices correctly within saturation. This problem is much less pronounced in STAG3, with the new saturation model presented in Chapter 4 giving much smoother transitions. Although there are still some deviations from the measured curves, the largest of these occur at high gate voltages, as explained above.

So far our evaluation of the STAG3 model has dealt only with body tied devices. Of course the model should also be able to accurately simulate floating body devices. The only experimental data for Technology A is for short channel devices, so we shall only evaluate for this case. This will provide the most challenging test of the model in any case, since the high lateral fields at the drain of a short channel device mean the impact ionisation has more of an influence on the characteristics.

Fig. 9.15 shows drain current plotted against drain voltage for 4 different gate voltages above threshold. While the fit is good in places, the main problem is that the onset of kink is not predicted accurately across the whole range of gate voltages. In this case, the optimal matching occurs around $V_{GfS} = 1.4V$. Above and below this value, the accuracy is gradually reduced, with deviations in prediction of the onset of up to $0.2-0.3V$. Fig 9.16 shows the drain conductance plot for the same device and operating conditions.

Part of the reason for the deviations may lie in the fact that the impact ionisation models

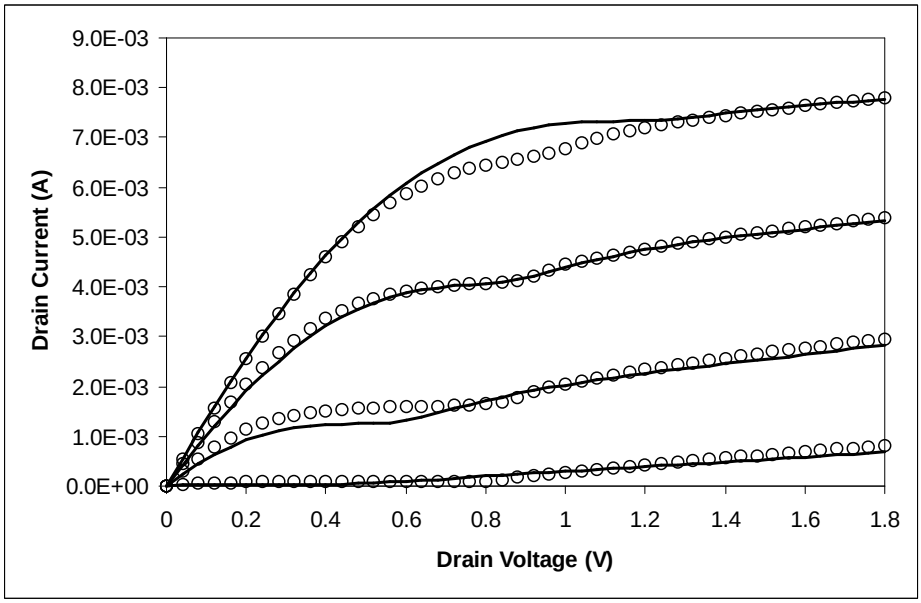

Figure 9.15: *Drain current versus drain voltage for $V_{GfS} = 0.6$, 1.0, 1.4, and $1.8V$ for a $W = 25\mu m$, $L = 0.25\mu m$ floating body device. Circles indicate measured values, solid lines indicate simulated results.*

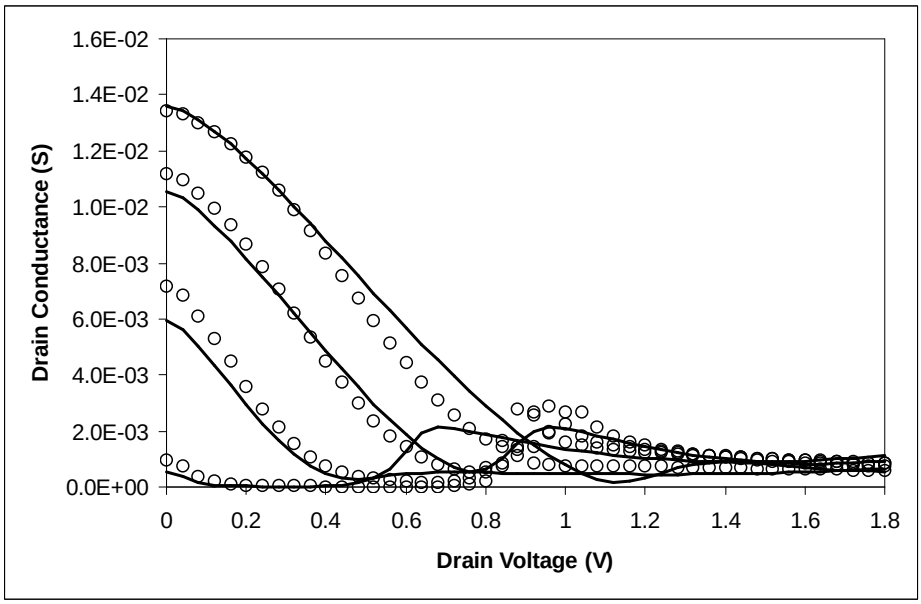

Figure 9.16: *Output conductance versus drain voltage for $V_{GfS} = 0.6$, 1.0, 1.4, and $1.8V$ for a $W = 25\mu m$, $L = 0.25\mu m$ floating body device. Circles indicate measured values, solid lines indicate simulated results.*

have not been changed since STAG2, and may not be suitable for accurate modelling of deep-submicron devices with high lateral fields. Another related issue is that as with all of the evaluation results, the model parameters have been hand fitted, rather than using computer optimisation. We have already seen from our body tied evaluations that this approach can give good matching with minimal effort for the core model. However, the impact ionisation model is a highly empirical auxilliary model with a large number of fitting parameters (7 if we include the temperature dependence). Not only is it difficult to judge appropriate values for parameters that lack a clear physical meaning, the task becomes progressively more difficult the more parameters are involved. So it may be that the existing model is adequate, provided that a software optimiser is used.

Overall, the new STAG3 model has been shown to provide good fitting across a range of channel lengths and biases, using just a single parameter set to obtain all results for the given process technology. Considering that all parameter optimisations performed for these simulations were simply hand adjustments, the level of accuracy is quite high. Full parameter optimisation would be expected to give better fitting, though some further model improvements are expected to be necessary for optimal experimental curve fitting.

References

[1] M S L Lee, *Compact Modelling of Partially Depleted Silicon-On-Insulator MOSFETs for Analogue Circuit Simulation*, PhD thesis, University of Southampton, Southampton, UK, Dec. 1997.

[2] R van Langvelde, *A Compact MOSFET Model for Distortion Analysis in Analog Circuit Design*, University Press Facilities, Eindhoven, 1998.

Parameter	Symbol	Description	Value	Unit
TNOM		Parameter measurement temperature	25	°C
TPG		Gate material type	1	-
TOF	t_{ox}	Front gate oxide thickness	4.5×10^{-9}	m
TB	t_{Si}	Silicon film thickness	90×10^{-9}	m
TOB	t_{box}	Back gate oxide thickness	380×10^{-9}	m
NSUB	N_B	Silicon film doping concentration	6×10^{17}	cm^{-3}
NP	N_P	Polysilicon gate doping concentration	1×10^{22}	cm^{-3}
NRSCE	N_{rsce}	Peak doping concentration at junctions	7.9×10^{17}	cm^{-3}
LRSCE	L_{rsce}	Characteristic length for lateral doping profile	6×10^{-8}	m
LD	l_d	Lateral diffusion	0	m
RSW	R_{sw}	Width scaled source series resistance	300	$\Omega.\mu$m
RDW	R_{dw}	Width scaled drain series resistance	300	$\Omega.\mu$m
VTEX	V_{Tex}	Extracted threshold voltage	0.64	V
VDEX	V_{Dex}	Drain-source voltage used to extract V_{Tex}	0.1	V
DELTA0	δ_0	Empirical threshold extraction parameter	0.07	-
SIGMA	σ	DIBL parameter	7×10^{-9}	m^{-1}
DELTAL	ΔL	Short channel effect parameter	2×10^{-8}	m
DELTAW	ΔW	Narrow width effect parameter	0	m
CHIFB	χ_{FB}	Temperature coefficient of flat-band voltage	9.5×10^{-4}	V/K
U0	μ_0	Low field carrier mobility	370	$cm^2/(V.s)$
ASR	α_{sr}	Surface roughness scattering coefficient	6×10^{-17}	$(m/V)^2$
APH	α_{ph}	Phonon scattering coefficient	2×10^{-4}	$(m/V)^{1/3}$
ACOU	α_{cou}	Coulomb scattering coefficient	0	m^3
VSAT	v_{sat}	Carrier saturation velocity	1.5×10^7	cm/s
K	k	Mobility temperature exponent	1.2	-
ALPHA0	α_0	Impact ionisation coefficient	5×10^6	m^{-1}
BETA0	β_0	Impact ionisation coefficient	2×10^6	V/m
LM	L_M	Ionisation length bias coefficient	4×10^{-8}	m
LM1	L_{M1}	Ionisation length bias coefficient	-7×10^{-9}	m/V
LM2	L_{M2}	Ionisation length bias coefficient	0	m/V^2
ETA	η	Impact ionisation empirical field adjustment parameter	0.5	-
LX	l_x	Short channel channel length modulation parameter	2×10^{-8}	m
VP	v_p	Short channel channel length modulation parameter	0.75	V
MEXP		Short channel channel length modulation exponent	4	-
JS	J_S	Diode diffusion current density	6.8×10^{-8}	A/m
ETAD	η_D	Diode diffusion ideality coefficient	1.23	-
JS1	J_{S1}	Diode recombination current density	9.6×10^{-8}	A/m
ETAD1	η_{D1}	Diode recombination ideality coefficient	1.65	-
FMIN	F_{min}	Minimum feature size	0.25	μm

Table 9.1: *STAG3 model parameters for simulation of Technology A device characteristics.*

Chapter 10

Conclusions and Further Work

10.1 Conclusions

The thesis began with a detailed treatment of the new low field surface potential model used in STAG3. It had been found that the old STAG2 model sometimes gave non-physical kinks in the transconductance curves around the onset into strong inversion when simulating deep sub-micron process technology. The model used by the Eindhoven model was adopted instead [1, 2], but this was found to give poor accuracy in subthreshold. In the end, the STAG3 model became a fusion of the STAG2 and Eindhoven models, with the additional feature that that possible presence of a lowly-doped polysilion gate is actually built into the core model.

Following that, new models for high vertical and lateral electric fields were derived. The old STAG2 mobility model, which had a linear dependence on vertical field, was replaced with a modern formulation which takes into account various scattering processes. An early form of this mobility model was presented in [3]; all the scattering terms are already present in their present forms, although the lateral field and saturation models have since undergone extensive improvement, and now include velocity overshoot.

The quantum mechanical model was then presented as part of the high field treatment. The method of combining the expressions for surface potential in strong inversion and saturation was demonstrated, which resulted in quantum effects automatically dropping to zero outside of strong inversion. This yielded a continuous one-piece high field expression for the STAG3 surface potential.

This was followed by a series of models dealing with shifts to the body factor or flat band voltage, such as non-uniform doping, SCE and RSCE, DIBL, and a new method for relating the STAG3 flat band model to extracted threshold voltages. This last sub-model was first published in [4].

Then the auxillary model was outlined, covering parasitics and floating body behaviour, self-heating and thermal behaviour, CLM, and external series resistance. This was fol-

lowed by the charge model, for which a new body charge expression had been devised for the drain region to eliminate a problem seen in STAG2. The relevant methodology, first published in [5], was also given.

Finally, some measured results were compared with simulations for body-tied MOS devices. It was found that a single parameter set could be used to get a good overall degree of fitting for a range of device sizes and terminal biases. The most noticeable problems occurred for higher drain biases, due to limitations in the charge expressions.

10.2 Suggestions for Further Work

10.2.1 DC Model

Many of the improvements which have led to the STAG3 model have centred around improving the accuracy of the DC model, in order to make it more applicable to devices in the deep sub-micron regime. Judging from the literature on the subject, we can expect STAG3 to yield good results for devices down to about 100nm gate length. However, as MOSFETs continue to be scaled down, certain physical effects become increasingly important, and these trends will need to be reflected in any future versions of STAG.

One feature currently missing from STAG3 is some way of modelling the gate current. This current results from quantum mechanical tunnelling of carriers across the potential barrier of the gate oxide and into the channel. The tunnelling probability increases exponentially with decreasing gate oxide thickness, so this effect becomes increasingly important as the front gate thickness is scaled down in modern CMOS processes [6].

Gate Induced Drain Leakage (GIDL) is the name given to a subthreshold phenomenon whereby a significant drain to bulk leakage current is induced, despite the drain voltage being well below the breakdown voltage [7]. As gate voltage is reduced, the subthreshold current starts to increase again. This occurs when the gate voltage becomes sufficiently small compared to the drain voltage, such that a strong drain-gate electric field is created. This strong transverse field allows valence electrons to tunnel into the conduction pair to form electron-hole pairs. These electrons and holes then flow into the drain and body respectively. This effect has even been observed in quarter micron technologies [8]. In order to accurately model this aspect of the subthreshold characteristics, a GIDL model needs to be added to STAG3.

One possible improvement has been identified for the mobility model, namely the inclusion of carrier inter-valley scattering. If a functional dependence on gate voltage (and drain voltage, if appropriate) could be identified, it could then be included as another scattering term in the Matthieson's rule expression. This would hopefully allow closer fitting across the full bias range.

Because we are concerned with simulating floating body SOI MOSFETs, accurate modelling of the body region is more important than for a bulk MOSFET model. It has been shown that body resistance in the form of a poor body tie can cause significant degradation of intrinsic device gain [9], and can even be worse than no body tie at all. More accurate modelling of the effect of a finite body resistance could be achieved by adding a number of internal bias-dependent resistances between the body and the source, drain, and external body nodes.

Also relevant to floating body simulations is the issue of the impact ionisation model. This has remained unchanged from STAG2, and further investigations are needed to determine whether the existing model provides sufficiently good fitting for short channel deep-submicron devices.

In the saturation region, the drain and source series resistances are no longer symmetrical, due to the presence of additional series resistance from the saturated drain region. If a value for this resistance could be calculated automatically, it would reduce the need to optimise the low field series resistance parameters to obtain a good fit at high drain voltages. The CLM model would provide a good starting point for this, since it already provides an expression for the length of the saturated drain region.

One numerical issue which remains in the STAG3 model is that the expression for the channel current is not symmetrical around the point $V_{ds} = 0$. For a symmetrical MOSFET device, we would expect identical behaviour in the forward and reverse modes (i.e. $I_{ds}(V_{ds}) = -I_{ds}(-V_{ds})$) It is shown in [10] that in order for this condition to be met, the second derivative of I_{ds} with respect to V_{ds} must be zero. In other words, when plotting the drain conductance g_{ds} as a function V_{ds}, the slope of this characteristic should be zero at $V_{ds} = 0$.

In order to overcome this problem, care must be taken when choosing the form of the velocity saturation expression [10], as well as the smoothing function used to model the transition between the linear and saturation operating regions [2]. Failure to do this can result in convergence problems moving from forward to reverse modes, and also renders the model unsuitable for performing distortion analysis. However, STAG3 has not shown any obvious convergence difficulties in this transitional region, and we are not concerned with developing a model for distortion analysis, so this issue was not given high priority.

A second numerical issue has been identified which relates to the linearisation of the body and channel charge expressions. It may be recalled that this is done to allow analytical integration of these expression to yield the total nodal charges. STAG3 follows the same general approach to STAG2; linearisation is performed around source surface potential. Linearising around the source is a common practise in compact models [11], but it does

introduce an asymmetry into the charge expressions, and as we saw in Chapter 9, it can lead to a reduction in model accuracy at high drain voltages. Furthermore, it results in channel charge not being allocated equally to the source and drain nodes when $V_{ds} = 0$.

The PSP model addresses this problem by linearising the charge expression around the surface potential mid-point [12]. This maintains the Gummel symmetry between source and drain. Unfortunately, this approach results in a linearised channel charge expression that has a functional dependence on the drain side surface potential. This complicates the equation for the channel current, and creates problems when differentiating with respect to the drain curface potential (which is how STAG calculates the saturation surface potential). Therefore, if the PSP methodology is to be adopted, it will necessitate other changes in the high lateral field model.

10.2.2 Charge and Noise Models

Another desirable improvement would be to further develop the body charge model discussed in Chapter 8. If detailed small-signal conductance data could be obtained for a range of channel lengths, it should be possible to gain a much clearer idea of how the body charge sharing varies with terminal bias, allowing a more complex and accurate model to be constructed.

One other aspect of the capacitance model which would benefit from improvement is the extrinsic p-n junction depletion capacitance model. Both STAG2 and STAG3 use the standard SPICE depletion capacitance model. However, there are a couple of problems with this standard model. The first problem is that we would expect the capacitance to saturate at some maximum value under forward bias, whereas the model does not provide any such limit. The second problem is that it is a piece-wise model which has discontinuities in high-order derivatives. Not only can this result in smaller time-steps being needed during transient simulations, but it can affect the accuracy of distortion analyses. A model to solve these problems was first proposed in [13]; not only does it use a single-piece, C∞-continuous expression, but it handles the dependence of the depletion capacitance on forward bias in a more physical manner. This work would provide a good starting point for improving this aspect of the STAG3 model.

The STAG3 model only provides a very rudimentary noise model. While it retains the standard expressions for the thermal and flicker noise that were present in STAG2, no further noise modelling capabilities have been added to this latest version. Given the advances made in models developed for the latest generation of compact circuit simulators [14], this is certainly one area where STAG could benefit from an update.

10.2.3 Other Device Models

One key feature which is currently unavailable in STAG is the facility to model fully-depleted device behaviour. Fully-depleted devices offer several advantages in performance over partially-depleted devices; these include improved sub-threshold characteristics, and reduction or elimination of electrical floating-body effects [15]. A number of fully-depleted models are already in existence [16, 17].

The viability of a model which can model transitions between partially-depleted and fully-depleted modes of operation has yet to be convincingly demonstrated. An older model does include this feature [17], but it employs an implicit equation, and as a result must be solved using an iterative procedure. It is stated in [17] that this iteration leads to an increase in CPU time of 30-40% when compared with closed-form analytical models. This is a very significant increase if the compact model is being used to model large circuits. An attempt to combine the two regimes was also made taking the BSIM3SOI model as a starting point [18]. This model was released into the public domain, but subsequent versions of BSIM3SOI reverted back to a dual model format, with entirely separate code being used to handle PD and FD simulations [19, 20].

While the ability to model fully-partially depleted switching is a desirable feature, there is little justification for including it in STAG it if it adversely affects model performance to any significant degree. Furthermore, time constraints meant that a full model implementation was not possible. It is nevertheless interesting to briefly consider some of the issues involved in constructing such a model.

Perhaps the most difficult problem from an implementation point of view will be to successfully determine the point at which switch-over from partially-depleted to fully-depleted operation occurs, and to express mathematically the degree to which any residual floating-body phenomena affect the operation of the device. It seems unlikely that basic depletion region theory will be sufficiently accurate to make such predictions, especially for short-channel devices, where the junction depletion regions will have a major influence. Empirical adjustments will undoubtedly be necessary, although ideally some deductions about the optimisation ranges might be made from examination of experimental data.

On the subject of experimental data, ideally we would like to evaluate the model against devices which actually exhibit mixed partial-full depeletion behaviour. Fortunately, such device data is available. At present, devices which exhibit such duality are more problematic to circuit designers than even partially-depleted devices. Most SOI compact models handle only one case or the other, and without a single unified model, the behaviour of these devices is unpredictable. Devices which are mainly fully-depleted, with residual partially-depleted behaviour near threshold, are more manageable, since the unpredictability is confined to a relatively small region of device operation. However, dual mode devices are ideal for testing a mixed model such as the one being proposed, since the characteris-

tics show a clear transition region between fully- and partially-depleted behaviour.

Another useful development would be to create a bulk MOSFET version of STAG. This would be relatively simple, and would require removal of the self-heating model (or at least have it switched off by default), and the removal of certain parasitic elements connected with the floating body. Additionally, the charge model would need to be changed to reflect the fact that the silicon body would be electrically connected to the substrate.

References

[1] Ronald van Langevelde, and Francois M. Klaasen, "Effect of Gate-Field Dependent Mobility Degradation on Distortion Analysis in MOSFET's", *IEEE Transactions on Electron Devices*, vol. 44, no. 11, pp. 2044–2052, 1997.

[2] R van Langvelde, *A Compact MOSFET Model for Distortion Analysis in Analog Circuit Design*, University Press Facilities, Eindhoven, 1998.

[3] J. Benson, N. V. D'Halleweyn, K. Mistry, and W.Redman-White, "A Physical Compact MOSFET Mobility Model Including Accuracte Calculation of Saturation Surface Potential", *Technical Proceedings of the 2003 NSTI Nanotechnology Conference and Trade Show,*, vol. 2, pp. 230–233, Feb. 2003.

[4] J. Benson, N. V. D'Halleweyn, W.Redman-White, C. A. Easson, M. J. Uren, O. Faynot, and J. L. Pelloie, "A Physically Based Relation Between Extracted Threshold Voltage and Surface Potential Flat-Band Voltage for MOSFET Compact Modeling", *IEEE Transactions on Electron Devices*, vol. 48, no. 5, pp. 1019–1021, May 2001.

[5] J. Benson, W.Redman-White, N. V. D'Halleweyn, C. A. Easson, and M. J. Uren, "Body Charge Modelling for Accurate Simulation of Small-Signal Behaviour in Floating Body SOI", *Solid State Electonics*, vol. 46, no. 4, pp. 529–537, 2002.

[6] H Watanabe, K Matsuzawa, and S Takagi, "Scaling effects on gate leakage current", *IEEE Transactions on Electron Devices*, vol. 50, no. 8, pp. 1779–1784, Aug. 2003.

[7] J Chen, T Y Chan, I C Chen, P K Ko, and C Hu, "Subbreakdown drain leakage in MOSFETs", *IEEE Electron Device Letters*, vol. EDL-8, pp. 515–517, 1987.

[8] J Chen, F Assaderaghi, P K Ko, and C Hu, "The enhancement of gate-induced-drain-leakage (GIDL) current in short-channel SOI MOSFET and its application in measuring lateral bipolar current gain β", *IEEE Electron Device Letters*, vol. 13, no. 11, pp. 572–574, 1992.

[9] C F Edwards, B M Tenbroek, M S L Lee, W Redman-White, and M J Uren, "The effect of body contact series resistance on SOI CMOS amplifier stages", *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2290–2294, Dec. 1997.

[10] K Joardar, K K Gullapalli, C C McAndrew, M E Burnham, and A Wild, "An improved MOSFET model for circuit simulation", *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 797–801, Apr. 1998.

[11] N Arora, *MOSFET Models for VLSI Circuit Simulation*, Springer-Verlag, 1993.

[12] T. -L. Chen and G. Gildenblat, "Symmetric bulk charge linearization of charge-sheet MOSFET model", *Electronics Letters*, vol. 37, no. 12, pp. 791–793, Jun. 2001.

[13] Colin C. McAndrew, Bijan K. Bhattacharyya, and Omar Wing, "A C∞-Continuous Depletion Capacitance Model", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, no. 6, pp. 825–828, Apr. 1993.

[14] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. J. Scholten, and D. B. M. Klaassen, "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation", *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1979–1993, Sep. 2006.

[15] J-P Colinge, *Silicon-On-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers, 1991.

[16] D. Suh and J. G. Fossum, "A physical charge-based model for non-fully depleted SOI MOSFETs and its use in assessing floating-body effects in SOI CMOS circuits", *IEEE Transactions on Electron Devices*, vol. 42, no. 4, pp. 728–737, Apr. 1995.

[17] J. W. Sleight, and R. Rios, "A Continuous Compact MOSFET Model for Fully- and Partially-Depleted SOI Devices", *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 821–825, 1998.

[18] Dennis Sinitsky, Stephen Tang, Arun Jangity, Fariborz Assaderaghi, Ghavan Shahidi, and Chenming Hu, "Simulation of SOI Devices and Circuits Using BSIM3SOI", *IEEE Electron Device Letters*, vol. 19, no. 9, pp. 323–325, 1998.

[19] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CA94720, *BSIMPD2.2 MOSFET Model User's Manual*, 2001.

[20] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley CA94720, *BSIMSOI v2.1 MOSFET Model - Users's Manual for BSIMFD2.1*, 1999.

Appendix A

Technology Key

The evaluation results presented in Chapters 6, 8, and 9 make references to Technology A and Technology B. These are the two PD-SOI technologies that have been used to evaluate the STAG3 model. Table A.1 provides values for the minimum gate length and nominal gate oxide thickness for these two technologies.

Technology	Minimum Gate Length (μm)	Gate Thickness (nm)
A	0.25	4.5
B	0.35	8

Table A.1: *Minimum gate length and nominal gate oxide thickness for Technologies A and B*

Appendix B

Publication List

- N. D'Halleweyn, **J. Benson**, W. Redman-White, K. Mistry, M. Swanenberg, *MOOSE: A physically based compact DC model of SOI LDMOSFETs for analogue circuit simulation*, IEEE Journal of Computer-Aided Design of Integrated Circuit and Systems, vol. 23 (10), pp. 1399-1410, Oct. 2004.
- K. Mistry, W. Redman-White, **J. Benson**, and N. D'Halleweyn, *A high speed dual modulus divider in SOI CMOS with stacked current steering phase selection architecture*, Radio Frequency IC Symposium, Philadelphia, USA, Jun. 2003, pp. 471-474.
- **J. Benson**, N. D'Halleweyn, K. Mistry, and W. Redman-White, *A physical compact MOSFET mobility model including accurate calculation of saturation surface potential*, Technical Proceedings of the 2003 Nanotechnology Conference and Trade Show, San Francisco, USA, Feb. 2003, vol. 2, pp 230-233.
- **J. Benson**, W. Redman-White, N. D'Halleweyn, C. Easson, M. Uren, *Body charge modelling for accurate simulation of small signal behaviour in floating body SOI*, Solid State Electronics, vol. 46 (4), pp. 529-537, Apr. 2002.
- N. D'Halleweyn, **J. Benson**, A. Aerts, M. Swanenberg, K. Mistry, and W. Redman-White, *SOI LDMOS modelling issues*, AUTOMACS Workshop: Modelling and Electrical Characterization of Advanced High-Voltage MOS Transistors, ESSDERC, Sep. 2001, Nuremberg, Germany.
- K. Mistry, W. Redman-White, **J. Benson**, and N. D'Halleweyn, *Low power multigigahertz divider architectures in SOI CMOS employing series current reuse*, European Conference on Circuit Theory and Design Proceedings, Helsinki, Finland, Aug. 2001, pp. 111.161-111.164.
- N. D'Halleweyn, L. Tiemeijer, **J. Benson**, W. Redman-White, *Charge model for SOI LDMOST with lateral doping gradient*, Proceedings of the 13th International Symposium on Power Semiconductor Devices and ICs, ISPSD '01, Jun. 2001, Osaka, Japan.
- **J. Benson**, N. D'Halleweyn, W. Redman-White, C. Easson, M. Uren, O. Faynot, and J.-L. Pelloie, *A physically based relation between extracted threshold voltage and*

surface potential flat band voltage for MOSFET compact modelling, IEEE Transactions on Electron Devices, vol. 48 (5), pp. 1019-1021, May 2001.

- N. D'Halleweyn, **J. Benson**, M. Swanenberg, and W. Redman-White, *A compact model for SOI LDMOST, including accumulation, lateral doping gradient, and high side behaviour*, 199th Meeting of the Electrochemical Society, Mar. 2001, Washington, USA.

- M. Lee, B. Tenbroek, W. Redman-White, **J. Benson**, M. Uren, *A physically based compact model of partially depleted SOI MOSFETs for analogue circuit simulation*, IEEE Journal of Solid-State Circuits, vol. 36 (1), pp. 110-121, Jan. 2001.